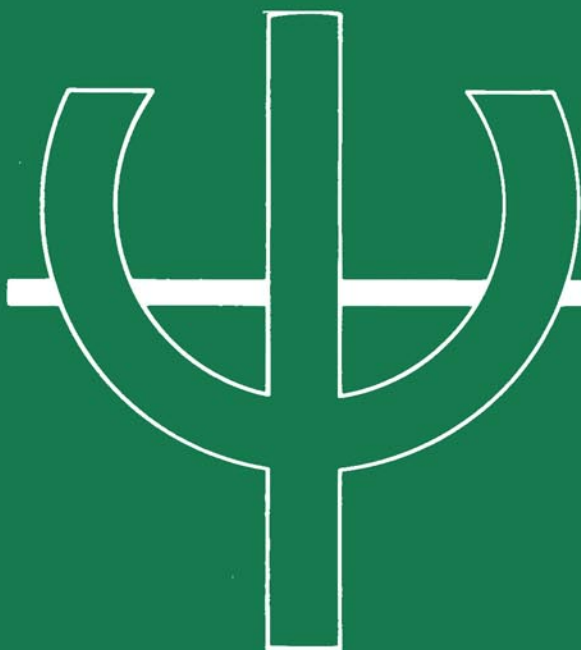




ADVANCES
IN
PSYCHOLOGY

16

Decision Making under Uncertainty

Edited by
Roland W. Scholz

North-Holland

DECISION MAKING UNDER UNCERTAINTY

ADVANCES
IN
PSYCHOLOGY
16

Editors

G. E. STELMACH

P. A. VROON



NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO

DECISION MAKING UNDER UNCERTAINTY

Cognitive Decision Research
Social Interaction
Development and Epistemology

Edited by

Roland W. SCHOLZ

IDM

*University of Bielefeld
FRGermany*



1983

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO

© ELSEVIER SCIENCE PUBLISHERS B.V., 1983

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0 444 86738 4

First edition: 1983

Second edition: 1987

Publishers:

ELSEVIER SCIENCE PUBLISHERS B.V.

P.O. Box 1991

1000 BZ Amsterdam

The Netherlands

Sole distributors for the U.S.A. and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.

52 Vanderbilt Avenue

New York, N.Y. 10017

U.S.A.

PRINTED IN THE NETHERLANDS

PREFACE

The editor of this volume likes to ask the question "*What for whom*" when being confronted with any new text; and he feels that the reader would be interested in a brief answer to this question

What is to be found in this volume?

In this volume one will find the revised papers of an international symposium on Decision Making under Uncertainty which was held at Bielefeld University from November 10th to 12th, 1982. The participants at this conference, psychologists (working mainly in the field of cognitive and social psychology), economists, philosophers, mathematicians and mathematics educators were asked to present new findings or to review the state of research on fallacies, biases and the development of decision behavior. The book is organized in five chapters. The Introduction attempts to outline the conceptual framework and how the three sections Cognitive Decision Research, Social Interaction and Development and Epistemology are interrelated, in particular how the research into social interactions might be reconnected with the research into individual decision making, and how new fields, such as research on developmental questions (i.e. from mathematics education and epistemology), can be productively integrated. As already shown by the scope of disciplines represented and by the chapter headings, both the subjective cognitive and the normative task-analytical sides of decision making are covered by the various chapters. Within the fifth and final chapter (short-) Comments are collected which evaluate the impact of the contributions on decision research itself, and also on cognitive psychology, developmental psychology, social psychology, economic theory and the discipline of mathematics education.

Who should be interested in the contents?

The contents should be of interest to theoretical or empirical researchers in those disciplines that are concerned with the adequate modelling of decision making. Apart from the above mentioned (sub-)disciplines (i.e. cognitive, developmental and social psychology, economy, philosophy, mathematical statistics and probability, mathematics education) certain fields of medicine, law, educational and social sciences should also be mentioned, as fundamental theoretical issues pertaining to these fields are discussed. However, the subject matter "decision making under uncertainty", its *conceptualization*, *modelling*, and *improvement* does not make it possible to provide either a closed list of fields of interest, or even for the weighing of its importance to the various fields of research. This is due to the fact that *decision theorists* and *decision analysts* as well as *decisions* and *uncertainties* are to be found in many places.

Acknowledgements

Science has its social dimensions. These become particularly important when integrating competing approaches or when trying to establish linkages between previously separate branches. I think, that the frame and perspectives taken by this volume on decision research could not have been possible if the symposium on 'Decision Making under Uncertainty' at Bielefeld University had not taken place. This symposium was arranged by the Institut für Didaktik der Mathematik (IDM) in cooperation with the Zentrum für interdisziplinäre Forschung (ZiF) and was sponsored by the Deutsche Forschungsgemeinschaft (DFG). Without the existence and the technical equipment of these institutes and without a grant from the Science Foundation this volume and its perspectives could not have been produced. The author wishes to thank the technical and administrative staff of these institutes, especially A. Bentrup, M. Onitshabo, and C. Buschmeyer, who did most of the technical editorial work, for their assistance in the arrangement of the conference and production of this book.

CONTENTS

Preface	v
DECISION MAKING UNDER UNCERTAINTY: Cognitive Decision Research, Social Interaction, Development and Epistemology	
Introduction to Decision Making under Uncertainty: Biases, Fallacies, and the Development of Decision Making <i>Roland W. Scholz</i>	3
COGNITIVE DECISION RESEARCH	19
The Theoretical Status of Judgmental Heuristics <i>Thomas S. Wallsten</i>	21
The Base Rate Fallacy Controversy <i>Maya Bar-Hillel</i>	39
The Two Camps on Rationality <i>Helmut Jungermann</i>	63
The Role of Heuristics in Models of Decision <i>Rudolf Groner, Marina Groner & Walter F. Bischof</i>	87
On the Testability of the Availability Heuristic <i>Klaus Fiedler</i>	109
Problem Structuring Calculi and Levels of Knowledge Representation in Decision Making <i>Patrick Humphreys & Dina Berkeley</i>	121
Verbal vs. Numerical Processing of Subjective Probabilities <i>Alf C. Zimmer</i>	159

Subjective Expected Utility: A Model of Decision Making <i>Baruch Fischhoff, Bernard Goitein & Zur Shapira</i>	183
SOCIAL INTERACTION	209
Conditions of Violating Individual Rationality <i>Werner H. Tack</i>	211
Biases in Group Decision Making <i>Helmut W. Crott & Johannes A. Zuber</i>	229
The Dependence of Cognition and Behavior in 2x2 Matrice Games <i>Ulrich Schulz & Theo May</i>	253
On the Prominence Structure of the Decimal System <i>Wulf Albers & Gisela Albers</i>	271
DEVELOPMENT AND EPISTEMOLOGY	289
Conceptual and Theoretical Issues in Developmental Research on the Acquisition of the Probability Concept <i>Roland W. Scholz & Manfred Waller</i>	291
The Perception of Probability Situations in Pupils Aged 12 to 14 <i>Georges Glaeser</i>	313
The Psychology of Inference and the Teaching of Probability and Statistics: Two Sides of the Same Coin? <i>J. Michael Shaughnessy</i>	325
(Mis-) Interpretation of Stochastic Models <i>Georg Schrager</i>	351
Randomness and Stochastic Independence - On the Relationship Between Intuitive Notion and Math- ematical Definition <i>Gerd v. Harten & Heinz Steinbring</i>	363
Decision-Theoretic Paradoxes <i>Georg Pflug</i>	375
COMMENTS	385
The Possible and the Profitable Some remarks on the symposium's impact on decision research <i>Baruch Fischhoff</i>	387

The Importance of the Fictional for Theories of Knowledge Some remarks on the symposium's impact on cognitive psychology <i>Alf C. Zimmer</i>	391
Why are Developmental Psychologists Biased with Regard to the Biasedness of the Human Subject? Some remarks on the symposium's impact on developmental psychology <i>Manfred Waller</i>	397
Individual Decision Making and Social Psychology Some remarks on the symposium's impact on social psychology <i>Katrin Borchering</i>	403
Towards a Theory of Limited Rationality Some remarks on the symposium's impact on economic theory <i>Reinhard Selten</i>	409
Towards an Education for Decision Behavior Some remarks on the symposium's impact on mathematical education <i>Hans-Georg Steiner</i>	413
ADDRESSES OF CONTRIBUTORS	421
ABSTRACTS	425
AUTHOR INDEX	437

This Page Intentionally Left Blank

INTRODUCTION

This Page Intentionally Left Blank

INTRODUCTION
TO
DECISION MAKING UNDER UNCERTAINTY:
BIASES, FALLACIES, AND THE DEVELOPMENT
OF DECISION MAKING

ROLAND W. SCHOLZ

Universität Bielefeld
IDM
F.R. Germany

Decision Theory and Research are genuinely interdisciplinary. Thus, it is impossible to identify either a single field of study or an individual scientific discipline which could serve to sufficiently explain the genesis of decision theory. While it is often stated that the roots of decision theory are to be found in mathematics and economy, the approaches, in these disciplines, have historically always been closely linked with the debate on rationality in philosophy. These disciplines have contributed central concepts of classical normative decision theory such as probability, utility, and heuristics. Decision theory, however, has also borrowed central concepts from the political and social sciences' mode of thought, such as, for instance, the concept of strategy.

The interest in analyzing human decision behavior has considerably increased in recent years. Statements like that of MARTIN SHUBIK: "One of the most important problems of our world is the trembling hand" (1982) have, in my opinion, dramatically pointed out the importance of the human factor, and especially of the consequences that are connected with the fuzzyness of human action. However, this fuzzyness is not restricted to motoric actions, but at least according to one "camp" of decision researchers (see JUNGEMANN, this volume) is also inherent in cognitive operations such as memorizing, or concept formation etc.. Within such fields as technology assessment, the possibilities and limits of the ability of man to decide sometimes create

sources of uncertainty that may even threaten our very existence. Traditionally decisions under uncertainty are defined by incomplete information or knowledge about a situation, i.e. the possible alternatives, or the probability of their occurrence, or their outcomes, are not known by the subjects. The sources of uncertainty are not homogeneous. Uncertainty may be internally ("I am not sure whether I have left my umbrella at John's house") or externally attributed ("It is not certain whether it will stop raining by tomorrow morning"). Different variants of these two types of uncertainty may be distinguished (cf. KAHNEMAN & TVERSKY 1982a) and further types may be defined when one considers the process of decision making (cf. BERKELEY and HUMPHREYS, 1982; HOGARTH, 1975).

During the last ten years, the scope of the "theory of decisions" has been expanded by the development of several psychological approaches in human decision behaviour research. The theoretical framework has been significantly extended by the elaboration of judgemental heuristics by DANIEL KAHNEMAN and AMOS TVERSKY. They took the fundamental ideas of HERBERT SIMON'S treatment of heuristics contained in his outline of a theory of bounded or limited rationality. This extension shows in fact that some kind of *cognitive decision science* which attempts to grasp the (cognitive) psychological foundations of decision behaviour, is beginning to evolve from the hitherto more quantitatively oriented *behavioral decision theory*. The theoretical approach of KAHNEMAN and TVERSKY was based on a series of experimental studies conducted during the seventies. While on the one hand these studies were widely, and sometimes enthusiastically received, on the other hand there has been pronounced criticism of this work. In particular, various objections have been voiced against the fallacy and bias argument which states that judgment and decision making show serious systematic errors, and the validity of the experimental results supporting these arguments has been criticized.

Although the fallacy concept is widely used it is hard to find a general definition of it, and there are obviously different comprehensions of what is meant by a fallacy. This is why I want to suggest the following definition of a fallacy before starting to outline the contents of the main chapters of this volume and describing how the different papers are interrelated.

A fallacy is the result of a cognitive process which, on the basis of information represented in memory, has led to a wrong conclusion, or to a wrong decision. A fallacy may consist of both the application of an inadequate model which has led to a deviating solution instead of a (possibly) existing definite formal solution, or of the application of rough rules of inference (such as intuitive estimates) used as a substitute, which systematically result in an inadequate, or incorrect solution. To me, this attempt to give a definition would seem to provide a satisfactory clarification of the fallacy or bias phenomenon which enables us to formulate a couple of questions and criteria with regard to it.

The core of this definition lies in the (externally evaluated, see SCHOLZ, 1981, p. 14 f.) incorrectness of the fallacy.¹⁾ This leads to the question: (1) How are the correct conclusions or inferences conceptualized? It should be noted, that according to this definition fallacies are not established by specific cognitive processes. All degrees of inferences which are inherent to knowledge application may be possibly judged to be fallacious, depending on the frame of reference or model of comparison (see below). Information represented in memory is the starting point for an inferential fallacy. The questions then are: (2) What information does an individual have represented in his mind with regard to a task he/she is to solve? (Under laboratory conditions of experimental study, this question can be formulated as follows: what structures are built up by subjects in the small world the experimenter has introduced? (cf. BERKELEY & HUMPHREYS, 1982).) (3) Which methods can be reasonably used in order to assess task understanding and the individual's level of information and knowledge? With regard to the first part of this question, I should like to note that the method that has dominated the literature on fallacies for a long time tends to assume that the subjects have understood the task in exactly the same way as the experimenter did.

Even more profound than the question pertaining to task representation is the problem of an adequate model to be used for comparison, i.e.: For which tasks can we be sure of having a generally accepted solution or model. In a very small number of cases we are able to draw on a model that is universally

¹⁾ I want to thank my colleague R. BROMME, Bielefeld, for his comments on the proposed fallacy definition.

accepted as in the instance of the following problem task, which I should like to quote as an example:

Example:

Imagine a circle the size of an orange. Now take a tape which is one metre longer than this circle's circumference, tie its ends together and place it so as to form a concentric circle around the smaller one. What will be the distance of this larger circle from the smaller, orange-sized one?

Now do the same, beginning with a circle having the circumference of the Earth. When a circle with a circumference that is one meter longer, is then placed concentrically around the earth sized circle, what is then the distance between both circles?

The solution or model for comparison, is in this case, the calculation of circle circumference. One is required to compare two radii of circles whose circumferences differ by one metre. It is a well-known fact of geometry that the difference between radii for a constant difference of circumferences is invariant with regard to the size of the smaller circle. The definiteness of the model for circle circumference, in my opinion, is due to the fact that the model for calculating the circle circumference for the object sought, is established as a mathematical theory, and a generally accepted theory of area calculus.

Leaving student errors in attempts at formal solution aside, subjects tackling this problem intuitively will generally assume that the distance between the two circles depends on the size of the smaller one.

Comparison of the circumference of a circle with its radius thus seems to be subject to a systematical proportion of error which involves a contradiction between intuition and intuitive solution on the one hand, and formal solution on the other. The contradiction in the given example is maintained (- at least for a while -) if the subject is confronted with the formally correct solution. Intuition, that is in this example the rough rule of inference which

says: "the bigger the circles, the less effect a constant difference of circumferences has on the distance between them" lets us down here.

The question now is whether this misleading intuition can be corrected. LIETZMANN, who discussed this problem in his 1917 problem collection under the title "Wo steckt der Fehler?" pointed out that there are, however, heuristics which can lead to a formal solution which is compatible with intuition. In a wording reminiscent of Socratic dialogue, he gives suggestions of the following kind; "When a hiker travels around the world, how much farther does his head travel, than his feet?" (LIETZMANN, 1969, p.8, translation by the author). He goes on to conclude that there will instantly be the feeling that there cannot be much of a difference. Other Socratic dialogues can be imagined such as first taking, as geometrical figures, straight lines of various length which are situated within rectangles with points of equal minimal distance, then continuing with squares and rectangles and other convex concentric polygons, ending up with considering the circle.

Even if insight as to the correct solution can be achieved in these ways, the typical tension between mathematical, formal-operative problem solving, and intuitive-concrete problem solving will be preserved or reappear after a period of time for many of the fallacies for which an "ideal mathematical model" is a comparative measure.

I should like to distinguish between fallacies of the circle task kind, and other fallacies. In the German language, there is the familiar distinction between Fehlschlüsse (fallacies) and Trugschlüsse (paralogisms). The latter is found in riddle mathematics. Paralogisms are something like problem formulations "in disguise". The following, rather sophisticated problem task will serve as an example (cf. WOLFF 1930, p. 124):

"An aeroplane requires one hour and twenty minutes to cover the distance Berlin-Hamburg. For the return flight from Hamburg to Berlin, the same aeroplane, travelling at exactly the same speed on exactly the same path, requires only eighty minutes.. Why?"

While a task analysis of the decoding process which confuses the cognitive structures, may be interesting from a linguistic point of view, such stories are unsuitable for the study of human errors and fallacies with the perspectives formulated by KAHNEMAN & TVERSKY (1982, p.494): "First, we expose some of our intellectual limitations and suggest ways of improving the quality of our thinking. Second, errors and biases often reveal the psychological processes and the heuristic procedures that govern judgment and inference. Third, mistakes and fallacies help the mapping of human intuitions by indicating which principles of statistics or logic are non-intuitive or counter-intuitive."

The "Cognitive Decision Research" Section

In the largest chapter "Cognitive Decision Research", some contributions to this volume attempt to define the position of psychological research with regard to *judgemental heuristics, biases, and fallacies*. In the literature on judgment and decisions, the base rate fallacy is most certainly the most extensively discussed phenomenon of human performance among those which are labelled as fallacies. The chapter about "THE BASE RATE FALLACY CONTROVERSY" (BAR HILLEL) summarizes the various psychological explanations given for this fallacy. While these considerations about fallacies cannot of course claim to be complete, they can be considered, in my opinion, to be representative in the way in which they treat the fundamental problems of the present research in psychological decision theory. Among these, the tasks of developing a broader psychological theory and of integrating various alternative approaches to explanation are notable. This problem is dealt with in the contribution "THE THEORETICAL STATUS OF JUDGMENTAL HEURISTICS"(WALLSTEN).

At present, no concept seems to be so frequently quoted in decision theory as the concept of heuristics.

Heuristics are generally understood to be rough cognitive rules of inference which reduce the complexity of situations. As principles of decision which reduce the complexity of situations, heuristics are conceived of as models of cognitive activity which make allowance for the operative and capacitive limitations of human cognitions. The status of heuristics (such as repre-

sentativeness heuristics or availability heuristics) as psychological explanatory constructs is not clear. The description of heuristics is often rather imprecise, and heuristics may for instance - as WALLSTEN (1980, p. 219) emphasizes, - partly be used arbitrarily as post hoc explanations of phenomena. Whether these shortcomings of research on heuristics can be overcome, depends, in my opinion, on two things:

First, it depends on whether heuristics can be defined in such a way that an empirically satisfactory test can be made. Second, it depends on whether the various heuristics that are treated more or less separately and side by side, can be integrated into a model encompassing all cognitive activity thus solving the elicitation problem of heuristics (cf. SCHOLZ, 1981, p. 80 ff).

Theoretical explanations for the deviation of human performance from normative models are connected with essentially different assumptions about *the nature of human rationality*. Within the paper on "THE TWO CAMPS ON RATIONALITY" (JUNGERMANN) a distinction is made between a pessimistic and an optimistic position. The pessimists consider judgmental biases, representational faults and coping effects/defects to be an essential part of human activity. The optimists argue that human performance is rational when regarding the cognitive costs of a decision (*meta rationality argument*), that the behavior is often well reasoned when decision behavior is conceptualized as a continuous process in time (*continuity argument*), or is not judged to be fallacious when the manner in which individuals have structured a problem (*structure argument*) is taken into account. In this context, the positions taken by the contributors to this volume are by no means uniform.

The concept of heuristic does not come from the theory of judgments and decisions. "THE ROLE OF HEURISTICS IN MODELS OF DECISION" (GRONER, GRONER & BISCHOF) studies the origins of this concepts within mathematics and philosophy while analyzing applications to the "Psychology of Thinking" and "Artificial Intelligence". In doing so, it clearly shows which specifications and theoretical integrations are required in order to subject this widely used concept to a satisfactory empirical test in the research field of judgment and decisions. These methodological problems are also elaborated in a study on judgments of subjects' contributions to various activities which inquires into

fundamental questions "ON THE TESTABILITY OF THE AVAILABILITY HEURISTIC" (FIEDLER).

Whether the aspiration to develop a cognitive decision theory out of a behavioral decision theory will be fulfilled depends mostly on the models of internal representations, on the memory models, and on the cognitive operations used in psychological models and theories. The contributions "PROBLEM STRUCTURING CALCULI AND LEVELS OF KNOWLEDGE REPRESENTATION IN DECISION MAKING" (HUMPHREYS AND BERKELEY) and "VERBAL VS. NUMERICAL PROCESSING OF SUBJECTIVE PROBABILITIES" (ZIMMER) present approaches which distinguish between different levels of abstraction or modes of task representation. Starting out from quantitatively different "decision making schemes" brings about, in my opinion, the possibility of more adequately grasping, theoretically, the broad interpersonal, but also intrapersonal variance of problem treatment.

The chapter on "Cognitive Decision Research" concludes with an extensive critical review of "SUBJECTIVE EXPECTED UTILITY (SEU)" (FISCHHOFF, GOITEIN, SHAPIRA). The SEU approach, together with simple linear models, has dominated modelling in decision theory for more than a decade without ever really questioning the validity of this approach as a model of cognitive structurization. This contribution reveals the theoretical shortcoming of the SEU-model but also takes the value of its application into account.

The "Social Interaction" Section

In social interactions, subjects also have to make individual decisions. Within the normative decision theory that is oriented towards formal modelling, it can be observed that since JOHN V. NEUMANN's and OSKAR MORGENSTERN's 1944 book "Theory of Games and Economic Behaviour", research into Individual Decision Making has separated itself from the experimental and theoretical research in game theory, which is concerned with "social interaction". This is remarkable insofar as the roots of modelling, that is the concept of utility, the concept of probability, and the economic principles of rationality such as maximizing profit and expectation, are identical in both fields, and that

fundamental deviations of behavior from these principles have been found not only in research into behavioral decision theory, but also in research into social interaction. Violating so-called individual rationality is one of these deviations. It occurs when a subject co-operates with others, obtaining less by this co-operation than he would obtain by not co-operating at all. The section "CONDITIONS OF VIOLATING RATIONALITY" (TACK) specifies conditions under which individuals violate this principle of rationality, permitting insights into the social structure of behavior and its determinedness, which is particularly dealt with in the section "THE DEPENDANCE OF COGNITION AND BEHAVIOR IN 2 x 2 MATRICE GAMES" (SCHULZ & MAY).

Understanding and modelling social interaction and group decisions will raise, even more pointedly than in the study of SEU validity the question whether a comprehensive model should be studied, or rather its components. The section on "BIASES IN GROUP DECISION MAKING" (CROTT & ZUBER) treats the phenomenon of the Group Choice Shifts frequently referred to as risky shifts, reporting three experimental studies. By means of these experiments, the authors show how far that which is possibly superficially interpreted as risky shift or group decision bias depends on the theoretically assumed (and possibly bargained for within groups) group decision rules, on assumptions about individual systems of preferences, and on the distribution of preferences within the group.

Early work on cognitive psychology (cf. MILLER 1956) has already pointed out the limitations of human information processing. Following the principles of bounded rationality (cf. SIMON 1955) in bargaining theory for instance, the aspiration level approach (see e.g. SAUERMAN & SELTEN 1962, TIETZ 1983, SCHOLZ 1980, ALBERS & ALBERS 1983) is based on the assumption that subject's bargaining behavior is governed by an aspiration scale which is given by the sets of bargaining alternatives. In the chapter "ON THE PROMINENCE STRUCTURE OF THE DECIMAL SYSTEM" (ALBERS & ALBERS) a model for the cognitive schemata operating on number systems is introduced which shows an essential resemblance to this approach. The model is illustrated and supported by empirical data from branded and unbranded department store articles. An evaluation of this model is given by applying it to four resp. five person bargaining games.

The "Development and Epistemology" Section

In my opinion, one of the essential deficits of decision research, is that it mostly conceives of human intuitions, cognitions, and the ability to decide as if they were static constructs. This is not only true in regard to the conceptualizing of a subject's decision behavior in a given situation (see above, and HUMPHREYS & BERKELEY, this volume) but also an invariance of fallacy is assumed in general human ability itself. TVERSKY's statement, reported by NISBETT & ROSS (1980, p. 14), "that there is no inferential failure that can be demonstrated with untrained undergraduates that cannot also (at least with a little ingenuity) be demonstrated in somewhat more subtle form in the highly trained scientist" is a very pronounced pessimistic statement formulating the developmental invariance of fallacies in decision making. Undoubtedly, many of the fallacies presently linked with the concept of probability, for instance, the TUROFF problem (cf. TUROFF, 1972, KAHNEMAN & TVERSKY, 1979, SCHOLZ, 1981 pp. 88), are arranged so as to create tensions even in persons familiar with the formal solutions, thus bringing about serious fallacies.

But: both the ontogenetic and the phylogenetic invariance of fallacies, however, would seem to be a fundamental misconception of the nature of cognitive activity, even if the latter is linked to discontinuities. The child's naive fallacy, termed artificialism, which says for instance that the rays of seeing go out from the eye is regarded by some developmental psychologists to be a universal phenomenon. To assume however, that the misconception associated with artificialism cannot be overcome, and that renowned physicists, for instance, fall back to the naive conception of the child in certain situations, is a very far-reaching and unsubstantiated proposition. Another example which could serve to demonstrate such a discontinuity is the conception of the world as a disc or a globe. Here again, as in the case of the 400 year-old concept of probability, distinct discontinuities can be found. These examples show, on the one hand, the developmental dependency of cognitive activity. On the other hand, they illustrate how far an educational and social-historical shaping of constitutive features of knowledge and thinking schemes influences cognitive activity.

Without any doubt the *concept of probability* is crucial for an analysis and a conceptualization of *uncertainty*. One should however take note that from a mathematical point of view there are several concepts and calculi of probability, and that one should distinguish mathematical from psychological objects. However, there are of course relationships between the individuals' intuitive naive knowledge and the body of formal scientific knowledge. This interrelationship is not only a problem of cognitive psychology but also an epistemological and educational one. Individual knowledge, for instance about the probability concept, depends on the interference of school and the cultural background. As LEVY-BRUHL (quoted according to PIAGET & INHELDER, 1975, p. XIV) pointed out, the absence of the idea of chance is an essential characteristic of the primitive mind. Today, there is a dispute among scientists as to which scope of situations and which type of uncertainty may reasonably be handled by the probability concept. For which situations a mathematical model based on probability may be meaningfully constructed is a matter of dispute among mathematicians and statisticians (cf. DINGES & ROST, 1982). Furthermore, psychologists have found out that the modes of coping with uncertainty in every day decision making are culturally dependent (cf. PHILLIPS & WRIGHT 1977, WRIGHT et. al. 1978).

The first paper presented in the chapter on "Development and Epistemology", "CONCEPTUAL AND THEORETICAL ISSUES IN DEVELOPMENTAL RESEARCH ON THE ACQUISITION OF THE PROBABILITY CONCEPT" (SCHULZ & WALLER), outlines the main variants of the philosophical interpretations of the probability concept and analyses how these concepts are to be dealt with in psychological research into the development of the probability concept. It is suggested that when both the objectivistic concept, and the intuitive component of probability leading to subjectivistic probability are jointly taken into account contradictions between different theoretical approaches in psychology with regard to the acquisition of the probability concept may be resolved.

In order to understand the calculus of probabilities used in simple urn experiments it is necessary to master more than fractional calculus. Also, the ability to solve simple probability tasks does not emerge spontaneously between the ages of 12 and 14. These are two of the results reported in the

contribution "THE PERCEPTION OF PROBABILITY SITUATIONS IN PUPILS AGED 12 TO 14" (GLAESER) which refers to the doctoral thesis of the Mexican researcher ALARCON.

The paper on "THE PSYCHOLOGY OF INFERENCE AND THE TEACHING OF PROBABILITY AND STATISTICS: TWO SIDES OF THE SAME COIN?" (SHAUGHNESSY) discusses conclusions which may be drawn from the theoretical and empirical research in psychology and mathematics that could lead to an improvement of decision behavior in the process of teaching probability and statistics in the classroom. Suggestions are made for both Psychology and Mathematics Education. Based on some empirical studies already concluded with college courses, a couple of concrete recommendations can be made such as: 1) "Teach in the following order from guesses, to experiments, to simulation, to theory." 2) "Emphasize the confounding role that belief systems can play in judgment and decision making." 3) "Sensitize students to misuses of statistics." 4) "Give students challenge problems which may require them to gather or simulate their own data, and then make a decision." 5) "Expose students to statistical paradoxes." 6) "... expose both teachers and students of probability and statistics to the results in the literature of judgment and decision making" (cf. SHAUGHNESSY, this volume).

The current courses in statistics and probability do not follow these ideas and obviously are not suitable to overcome the shortcomings that prevent an adequate modelling of random events. This is one of the propositions presented in the paper "(MIS-) INTERPRETATION OF STOCHASTIC MODELS" (SCHRAGE). Furthermore, it reports serious misapplications of probability calculus in a murder trial, medical diagnoses and even in a mathematical proof on Fermat's conjecture (an unsolved problem of number theory) that has been conducted and published by professional mathematicians, and that has also been received awards from an editorial committee (including university teachers).

The linkage of preconcepts and intuitions of probability on the one hand and mathematical calculi and techniques on the other hand is decisive for an instruction of probability in schools that would lead to more successful decision making. The contribution "RANDOMNESS AND STOCHASTIC INDEPENDENCE - ON THE RELATIONSHIP BETWEEN INTUITIVE NOTION AND MATHEMATICAL DEFINITION"

(v.HARTEN & STEINBRING) points out firstly, that the distinction between the theoretical concept of probability and its application is crucial for an adequate modelling and interpretation of stochastic events, and secondly, that there is no randomness in an absolute sense. What is judged to be stochastically independent depends on the body of techniques being used, and on the different mathematical and intuitive approaches to independence and randomness. The theoretical and conceptual frame of reference of decision theoretic modelling and with it the intuitive foundations are in a state of continuous change. That which is perceived to be paradoxical at one stage of scientific development or by one school of thought may be completely clear and undisputed from another point of view. Within the paper "DECISION - THEORETIC PARADOXES" (PFLUG) one can find a variety of examples partly taken from the history of statistics which have established or are still establishing severe deviations from widely accepted and followed normative criteria of statistics.

The impact of this volume on various disciplines

The topics discussed in this volume are of interest for several disciplines. The impact of the contributions presented in this volume on DECISION RESEARCH (FISCHHOFF), COGNITIVE PSYCHOLOGY (ZIMMER), DEVELOPMENTAL PSYCHOLOGY (WALLER), SOCIAL PSYCHOLOGY (BORCHERDING), ECONOMIC THEORY (SELTEN), and MATHEMATICS EDUCATION (STEINER) is outlined by researchers from these disciplines who were present at the Symposium. Of course, other disciplines, e.g. medicine, social sciences, mathematics, that were not represented, might also be affected or challenged by the results and propositions documented in this volume. Naturally, the representatives of the different disciplines emphasize different aspects. Although from a decision theoretical perspective, *methodological* problems and questions of research strategies (e.g. top down vs. bottom up) seem to be most significant, *conceptual* issues about the nature of human knowledge (for instance "THE IMPORTANCE OF THE FICTIONAL FOR THEORIES OF KNOWLEDGE", see ZIMMER's comment, or the concepts of limited rationality in economic theory, see SELTEN's comment) are regarded as creating important research problems in other fields. The comments made by the mathematics educational view show the innovative power of a growing discipline. The methodology of mathematization is shown to be inextricably connected with the social dimensions of learning and instruction. Mathematics in general (and not only statistics and probabil-

ity theory) is losing its unique feature of always being either right or wrong when put into a social context (e.g. the classroom). Furthermore, as several of the papers point out, if the dynamic views were to be emphasized, we may not only expect decision research to have an impact on mathematics research but also the other way round.

Looking ahead

When considering the outline of contents the question might arise as to whether it is at all possible to bring such different subjects together. The editor of this volume is rather optimistic in this respect. The arguments supporting this optimism are at least threefold. First, there is the necessity for a multitude of approaches due to the cognitive, subjective and to the normative ties of modelling and understanding decisions. These ties imply a methodological and disciplinary multitude. Secondly, broadening the frame need not necessarily lead to an increase of information, complexity and confusion. A theory building orientation increases the possibility of restructuring and reintegrating. Thirdly, some experiences with regard to the interaction within this "broad frame" were already gained during the Symposium about "Decision Making under Uncertainty". The enterprise of confronting research(ers) from economics with developmental psychology, or those from mathematics (education) with those from group decision making in social psychology for instance may seem to be adventurous but, as we found out at Bielefeld, it worked well, and seems to be very promising (e.g. with regard to the second argument) for future research in decision making.

References

- Albers, W. & Albers, G.: 1983. Prominence and aspiration adjustment in location games. In: Tietz, R. (Ed.) *Aspiration levels in Bargaining and Economic Decision Making*. Heidelberg: Springer, 243-259
- Berkeley, D. & Humphreys, P.C.: 1982. Structuring Decision Problems and the Bias Heuristic. In: *Acta Psychologica* 50, 201-252
- Dinges, H. & Rost, H.: 1982. *Prinzipien der Stochastik*. Stuttgart: Teubner
- Hogarth, R.M.: Cognitive Processes and the Assessment of Subjective Probability Distributions. In: *Journal of the American Statistical Association*, June 1975, Vol. 70, No. 350, 271-289

- Kahneman, D. & Tversky, A.: 1982a. Variants of uncertainty. In: Kahneman, D./ Slovic, P. & Tversky, A. (Eds.), *Judgement under uncertainty; Heuristics and biases*. Cambridge: Cambridge University Press, 509-520
- Kahneman, D. & Tversky, A.: 1982b. On the study of statistical intuitions. Kahneman, D./Slovic, P. & Tversky, A. (Eds.), *Judgement under uncertainty; Heuristics and biases*. Cambridge: Cambridge University Press, 493-508
- Lietzmann, W.: 1969 (1st ed. 1917). *Wo steckt der Fehler*. Leipzig:
- Miller, G.A.: 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. In: *Psychological Review* 63, 81-97
- Nisbett, R.E. & Ross, L.: 1980. *Human inference: Strategies and shortcomings of social judgment*. New York: Prentice-Hall, Englewood Cliffs
- Phillips, L.D. & Wright, G.N.: 1977. Cultural differences in viewing uncertainty and assessing probabilities. In: Jungermann, H. & deZeeuw, G. (Eds.), *Decision making an change in human affairs*. Amsterdam: D. Reidel, 507-519
- Piaget, J. & Inhelder, B.: 1975. *La genèse de l'idée de hasard chez l'enfant*. Paris 1956: Presses Universitaires de France. Translated as: *The Origin of the Idee of chance in Children*. New York 1975: Norton
- Sauermann, H. & Selten, R.: 1962: *Anspruchsanpassungstheorie der Unternehmung*. In: *Zeitschrift für die gesamte Staatswissenschaft* 118, 577-597
- Scholz, R.W.: 1980. *Dyadische Verhandlungen; Eine theoretische und experimentelle Untersuchung von Vorhersagemodellen*. Meisenheim am Glan: Hain
- Scholz, R.W.: 1981. *Stochastische Problemaufgaben: Analysen aus didaktischer und psychologischer Perspektive*. Bielefeld: Universität Bielefeld
- Shubik, M.: 1982. Personal communication during a lecture hold at Bielefeld University
- Simon, H.A.: 1955. *Models of Man*. New York: Wiley
- Tietz, R. (Ed.): 1983. *Aspiration levels in Bargaining and Economic Decision Making*. Berlin, Heidelberg, New York, Tokyo: Springer
- Turoff, M.: 1972. An alternative approach to cross-impact analysis. In: *Technological Forecasting and Social Change*, 309-339
- Tversky, A. & Kahneman, D.: 1979. Causal Schemas in Judgments Under Uncertainty. In: Fishbein, M. (Ed.), *Progress in Social Psychology*. Hillsdale: L.E.A., 49-72

- Wallsten, Th.S.: 1980. Processes and models to describe choice and inference behavior. In: Wallsten, Th.S. (Ed.): Cognitive processes in choice and decision behavior. Hillsdale: L.E.A.
- Wolff, Th.: 1981. Vom lachenden Denken. Ein Buch von Wundern und Problemen. Berlin: Scherl
- Wright, G.N./Phillips, L.D./Whalley, P.C./Choo, G.T.G./Ng., K.O./Tan, I. & Wisudha, A.: 1978. Cultural differences in probabilistic thinking. In: Journal of Cross-Cultural Psychology 9, 285-299

COGNITIVE DECISION RESEARCH

This Page Intentionally Left Blank

THE THEORETICAL STATUS OF JUDGMENTAL HEURISTICS¹⁾

THOMAS S. WALLSTEN

University of North Carolina
L.L. Thurstone Psychometric Laboratory
USA

The recent marked shift in empirical research on judgment from testing normative models to investigating people's biases and heuristics has been extremely valuable in many respects. However, it also has certain shortcomings, as would be expected with any major change in research focus. There are three interrelated problems: (i) an almost excessive concern with the three heuristics of representativeness, availability, and anchoring and adjustment; (ii) a tendency toward treating classification of heuristics as explanation; and (iii) a lack of rigorous theory development. The last is the most serious problem, of course, and there are recent, encouraging trends suggesting that it may be temporary.

The plan of this paper is to first discuss the problems just outlined. Then, having concluded that there is a serious need for theory development, I will provide a very selected review of the literature to summarize some of the findings that any complete theory must explain and to indicate limitations that I perceive with some of the current experimental work.

¹⁾ I am grateful to Austen Clark, Samuel Fillenbaum, and Barbara Forsyth for critically commenting on an earlier draft of this paper, and to Anita Bergmann, Gerda Wallsten, and Herman Wallsten for translating chapters by Scholz (1981).

The problems

The three heuristics of representativeness, availability, and anchoring and adjustment were, of course, first identified by Tversky and Kahneman (1971, 1973, 1974; Kahneman & Tversky, 1972a, 1973), and by now the literature on these heuristics is enormous. Their existence is probably the major claim that people outside of the field now know about judgment under uncertainty.

However, there are many other interesting features of judgment which, while they might not be termed heuristics, certainly provide insight into the judgment process. For example, there are commonly documented primacy and recency effects in judgment. In the former case, early information either causes subsequent information to be subjectively distorted (e.g. Anderson & Barrios, 1961; Asch, 1946; Wallsten, 1981), or causes the seeking only of information which supports or confirms already existing opinion (Mynatt, Doherty, & Tweney, 1977). In the latter case, subsequent information in a sequence exerts an undue influence on judgment (e.g. Anderson & Hubert, 1963). Another well-documented bias is that of value-induced distortions (Irwin, 1953; Irwin & Snodgrass, 1966; Delaney & Wallsten, 1977). In this case, positive and negative values of events, or outcomes, influence their judged probabilities, and cause information to be distorted for or against the events.

I am not proposing that six heuristics and biases ought to be investigated instead of three. Rather, the point is that research which focusses on any number of heuristics, simply asking the conditions under which each appears, stops short of some of the most interesting and useful questions. Classification is clearly very important, but there is the danger in such research, frequently succumbed to, that the description of the phenomenon is taken as its explanation, and underlying theoretical principles are not sought. A search for underlying principles rather than a fixation on disparate phenomena will lead to conclusions of greater generality and usefulness.

Related to the tendency towards explanation by classification is the considerable debate concerning whether or not certain judgments are rational (see e.g. Cohen, 1981, and the subsequent commentaries). Cohen (1979) has

pointed out that such evaluations assume independent definitions and consensus as to what is rational. The question of what constitutes rationality is clearly interesting and important, but my point here again is that the arguments on evaluation focus research away from underlying processes.

When the processes by which people encode and judge probabilistic information are sufficiently well understood, then it will be possible to predict the conditions under which individuals use specific heuristics or are subject to specific biases. Fortunately, theories along these lines are developing, although that work is still in its early stages and is quite incomplete. See, for example, Groner's chapter in this volume, Tversky and Kahneman's (1980) or Einhorn and Hogarth's (1982a, b) theories on judging causality, or my work (Wallsten 1976, 1977; Wallsten & Barton 1982) on sequential processing of differentially salient dimensions. Interestingly, both Einhorn and Hogarth's and my theories rely strongly on an anchoring and adjustment mechanism. To be clear, one does not desire a separate theory for each heuristic or bias, but rather a single theoretical framework to predict the range of judgmental effects that are observed.

Literature review

I turn now to a selected review of the literature to summarize some of the effects for which any complete theory must account, and to indicate problems with some experiments that become apparent when considering theoretical explanations. The research specifically testing availability or anchoring and adjustment is relatively sparse (but see Fiedler's chapter on availability in this volume). A possible reason for the relatively little empirical work and good agreement among results is that the conditions for eliciting these two heuristics appear to be straightforward and tied rather directly to well-defined aspects of the task. At least superficially, it is clear when the requisite conditions are present, and when they are not, the heuristics are not expected. Indeed, as already indicated, the anchoring and adjustment process is sufficiently well specified that it can serve as a basis for theory development.

Matters are different with the representativeness heuristic. Here there has

been a lot of research with a wide spectrum of results. There are at least three reasons for this state of affairs. First, representativeness has not been, and perhaps cannot be, defined a priori in terms of specific features or relations. Second, predictions from the representativeness heuristic are generally in terms of features of the stimulus that will not influence judgment. Thus, the predicted experimental outcome is frequently that of the null result. Third, many of the experiments have been conducted in the context of word problems. The phrasing and choice of words is crucial and can have a substantial impact on the way the information is encoded.

Base Rates

An example of the prediction of a null result concerns the use of base rate information. According to Kahneman and Tversky (1973), a person's judgment of the probability that event A belongs to class B, rather than to some other class, depends only on how representative A is of B, and not on the base rate of B in the population. A chapter by Bar-Hillel in this volume reviews the literature on use of base rates, and it is therefore not necessary to provide another detailed review here. However, the perspective in this chapter is different from hers, so some additional discussion is warranted.

First, recall the basic phenomenon as demonstrated by Kahneman and Tversky (1973). In one experiment subjects read brief personality sketches, and for each one rated the likelihood that the described person was a graduate student in each of nine fields. These ratings correlated relatively poorly with independently judged base rates, i.e., judged proportions of graduate students in each of the fields (mean $r=.17$), but correlated highly with independently rated similarities of each personality sketch to the stereotype graduate student in each field (mean $r=.91$). In another experiment subjects rated personality sketches according to whether the described individuals were more likely engineers or lawyers, knowing that each sketch came from a population consisting of 70 % engineers and 30 % lawyers (or the converse). Ratings varied with the personality sketches, but were independent of the population base rates.

Although base rates are clearly important in all Bayesian calculations, there are situations or modes of thinking within which their use is questionable,

as Cohen (1979), Shafer (1981), and a few others have pointed out. An example provided by Cohen (1979) is one in which 1,000 people are watching a rodeo; it is known that 400 of them have paid and assumed that the remainder have entered through a hole in the fence. Under those conditions, it would generally be agreed that an arbitrarily selected person should not be prosecuted for failure to pay simply because the probability that he failed to do so is .6. To expand the example, given a fixed amount of evidence against the person, he would not be considered to be more likely guilty when 700 people entered through the fence than when 600 people did.

Nevertheless, under many conditions base rates do provide useful information, regardless of the mode of thinking, and it is of interest to inquire into their role in judgment, particularly when the representativeness heuristic suggests that they will be ignored. In fact, there are various conditions under which base rates are attended to, and thus far at least three factors have been identified that affect their use. These are (i) expertise of the judge, (ii) relevance or specificity of the information, and (iii) salience of the information.

Expertise. There is evidence that experts working in their field do not demonstrate many of the same limitations shown by subjects in experimental situations. Thus, physicians appear to be sensitive to base rates when judging the probabilities of diseases. In a recent experiment (Wallsten 1981), experienced physicians and fourth year medical students rated likelihoods for 50 diseases based on initial information, then following historical and physical data, and then again following laboratory data, for each of a large number of hypothetical patients. The medical students were quite insensitive to base rates; they initially gave a very large number of .5 likelihood ratings, which then propagated throughout the examination. The physicians, on the other hand, used considerably fewer .5 estimates. Instead, they provided estimates that were much closer to zero, which then tended to decrease further as the examination continued, as was appropriate for the information they were receiving. Christensen-Szalanski and Bushyhead (1981) also found evidence in ongoing clinical situations that physicians were sensitive to base rates.

Specificity of information. Bar-Hillel (1980) suggested that base rates tend to be ignored when subjects do not perceive them to be relevant to the judgment task at hand. Without fully defining the concept of relevance, she suggested, correctly I think, that the smaller the subset of individuals or events to which the information refers, the more relevant it is deemed to the judgment at hand. Thus, for example, to learn that a given individual enters through a hole in the fence 60 % of the time that he goes to an event is more relevant than to learn that 60 % of the people at the event have snuck in.

Using a series of problems based on Kahneman and Tversky's (1972b) taxi cab problem, Bar-Hillel (1980) manipulated the subset of the population to which the base rate or diagnostic information referred. She either lowered the relevance of the diagnostic information by making it apply less specifically to the event in question, or increased the relevance of the base rate by making it apply more specifically to the event. In these cases, subjects' judgments represented an integration of base rates with diagnostic information. In an earlier experiment using Tversky and Kahneman's (1973) lawyer/engineer paradigm, Carroll and Siegler (1977) demonstrated that subjects do use base rates, but do so to a greater degree when the base rate refers to a small sample, each member of which is to be judged, than when it refers to a large population, only a small fraction of which is to be judged. This can be thought of as another way to increase the specificity, or relevance of base rates.

Salience of base rates. Nisbett, Borgida, Crandall, and Reed (1976) have suggested that base rate information tends to be ignored because it is "remote, pallid, and abstract", in contrast to diagnostic information, which is "vivid, salient, and concrete". This hypothesis was supported in an experiment by Borgida and Nisbett (1977) in which prospective psychology majors rated psychology courses based either on catalog information only, or on that information supplemented by mean student course evaluations or by face-to-face comments from older students. The "pallid" course evaluations had little impact on ratings, while the "vivid" comments had substantial effect.

Indeed, it may be that a major reason why base rates have been ineffective

in experiments is because they are stated once, and then allowed to fade into the background while diagnostic or target information varies over trials. Consistent with this suggestion, Fischhoff, Slovic, and Lichtenstein (1979) demonstrated that subjects are very responsive to base rates when the base rates vary from trial to trial.

Base rates may also be more salient if they are actually experienced over time, rather than simply stated in summary form. This was demonstrated in four experiments by Manis, Dovalina, Avis, and Cardoze (1980), in which subjects experienced base rates by seeing a sequence of photographs of some individuals in the population while being told the relevant behavior or belief of each one. These base rates then affected subjects' judgments of the behavior or beliefs of other members of the population. Manis et al. also attempted to manipulate concreteness of the base rates through the format in which they were presented, but found this manipulation to add nothing beyond the effect of experience itself.

Christensen-Szalanski and Beach (1982) also gave subjects experience with base rates, but then, in a switch from other experiments, manipulated whether diagnostic information was simultaneously experienced with the base rates or was subsequently presented in summary statistical form. Interestingly, only the subjects simultaneously experiencing both types of information integrated them when later judging the posterior probability of an event. Judgments by the other subjects reflected only the summary diagnostic data and not the experienced base rates. Also, 10 % of this latter group failed to respond, indicating they did not understand or could not answer the question, whereas none of the former group failed to respond. Christensen-Szalanski and Beach (1982) suggest that base rates are more readily used when they are experienced in relation to diagnostic data. If this conclusion is supported in future research, then simple notions of salience will not explain under utilization of base rates.

In another study attempting to make base rates more concrete, Carroll and Siegler (1977) tested the hypothesis that subjects were more likely to use base rates if they could be translated into concrete numbers of people. One group of subjects judged 20 personality sketches from a population with a

70:30 split, which could easily be translated into 14 lawyers and 6 engineers, or the opposite; while another group had a 75:25 base rate which could not be translated into an exact number of lawyers or engineers. In two experiments, one with neutral and one with informative sketches, base rate had a significant effect. Translatability of base rates had an interactive effect with the neutral sketches only. In that case, base rate was more influential in the translatable 70 % condition than in the non-translatable 75 % condition.

To summarize the research on use of base rates, there is no doubt that in many problems base rates are either ignored or insufficiently weighted. However, the generalization that subjects judge according to representativeness and therefore ignore base rates is clearly overstated. At least three factors affect the use of base rates: subject expertise, information specificity, and information concreteness or salience. Furthermore, there is the additional result by Christensen-Szalanski and Beach (1982) that 10 % of the subjects did not understand the problem when parts of it were presented in summary form. Problem comprehension has not generally been assessed, but it may be an important factor in determining which information is utilized and which is ignored.

Sampling distributions

The representativeness prediction regarding judgments of sampling distributions is also a negative one. On the assumption that when people use sample statistics to estimate population parameters, they judge the similarity of the statistic to the parameter, it is predicted that judgments will be insensitive to sample size. Kahneman and Tversky (1972a) provide many examples of this heuristic at work, two of which will be summarized here. In one experiment, subjective sampling distributions were estimated by asking large numbers of subjects questions of the following sort: On what percentage of days will the number of boys among 1000 babies born be as follows: up to 50 boys, 50-150 boys, ..., 850-950 boys, more than 950 boys? Problems differed in terms of the size of the sample, being either 10, 100, or 1000 babies in the above instance, and also in terms of the particular variable over which the distribution was to be estimated. In all cases the subjective sampling

distribution was independent of sample size, a "universal sampling distribution", in Kahneman and Tversky's words.

In another experiment utilizing three different problems, some subjects were required to judge whether a large sample or a small sample was more likely to yield an extreme outcome, or whether the probability was about equal for the two samples. Other subjects made the judgment with regard to a non-extreme outcome. For each outcome, the majority of the subjects judged it to be equally likely in the two samples, thereby ignoring the impact of sample size.

Studies looking into the generality of these results have indicated that sample size is not always ignored, and have identified at least three factors that affect judgments about samples: (i) individual differences; (ii) sample to population ratio; and (iii) the wording of the problems.

Individual differences. In commenting on the Kahneman and Tversky (1972a) results, Evans and Dusoir (1977) suggest that "the construction of their problems seems unnecessarily complex and ... their tests are not designed to look at individuals, thus failing to provide a powerful test of the hypothesis that the role of sample size is understood by some but not all subjects" (p. 130). In one experiment Evans and Dusoir (1977) used binomial samples that varied in sample proportion as well as sample size. Subjects were shown all possible pairs of the binomial samples, and in each case were required to state which sample gave better evidence of bias in the underlying population. In this context, individual subject analyses showed that a clear minority of the subjects (11 of 48) were responsive to sample size considerations, while the remainder were generally insensitive to sample size.

Recently, Krantz, Nisbett, and their associates have also demonstrated strong individual differences in the nature of statistical reasoning. In one study (Jepson, Krantz, & Nisbett 1982) subjects were given a series of word problems to solve, where the correct solution required recognition of either the importance of sample size or the principle of regression to the mean. For each problem, subjects provided brief descriptions of how they arrived at their solution. Each description was then scored as statistical if in some

fashion it mentioned the correct principle, or as deterministic or intermediate otherwise, according to certain criteria. For our purposes the important result was a significant correlation indicating that subjects with higher SAT's tended to give more statistical responses. In another study (Krantz, Fong, & Nisbett 1982) the tendency to provide statistical descriptions increased dramatically with degree of statistical training. Furthermore, the use of statistical reasoning was greater if the wording of the problem specifically indicated a random force at work.

Sample to population ratio. Bar-Hillel (1979) suggested that subjects frequently appear to be sensitive to sample size as an artifact of responding to perceived sample to population ratio, rather than because they are responding to sample size, per se. To test this notion, she told subjects about nine hypothetical voting surveys which differed in terms of the number of people in the survey sample and the number of people in the voting population. Subjects had to judge the accuracy of each survey. As predicted, they were most responsive to the sample to population ratio, but interestingly, within ratios they were responsive to sample size. Normatively, of course, when sampling from a finite population the sample to population ratio is an important consideration. However, given the population sizes used in these problems, sample size alone should have been the primary, not the secondary, determinant of confidence.

Problem wording. Returning to Evans and Dusoier (1977), they thought that in Kahneman and Tversky's (1972a) original experiment subjects may have appeared indifferent to sample size, because the problems were worded in a confusing manner. Thus, for example, in the original experiment, one problem described a large and a small hospital, and then stated: "For a period of one year each hospital recorded the days on which more than 60 % of the babies born were boys. Which hospital do you think recorded more such days?" Evans and Dusoier thought that paragraph might be complicated with regard to two concepts: "most days included in one year", and "more than 60 % of the babies". Thus, they gave the original problem to one group of subjects but then changed the problem for another group of subjects to read: "Which hospital do you think is more likely to find on one such day that all the babies born were boys?" Two other groups of subjects received problems that were simplified with re-

gard to only one or the other of the two concepts. Of the 20 subjects receiving each problem, 11 provided the correct answer for the problem that exactly duplicated Kahneman and Tversky's, while 48 of the remaining 60 subjects gave the correct answer to simplified problems. In a different context, Bar-Hillel (1979) found a similar result.

Olson (1976) also questioned the generality of the Kahneman and Tversky results. He designed three experiments on the assumption that subjects were responding to concrete features determined by the problem wording, rather than to abstract representations of the task. Thus, for example, regarding the universal sampling distribution, Olson thought that perhaps subjects were responding simply to the ordinal positions of the categories rather than to the degree that a sample characteristic represented a population characteristic. To test this hypothesis, he obtained subjective sampling distributions by having subjects estimate the proportion of a sample that would be expected in each of 11 categories. Both sample size and category size varied over subjects, but the percentage of the sample put in each category was the same in all cases, thereby supporting Olson's hypothesis.

Two additional experiments inquired into the features that govern whether a sample is judged representative of a population. Kahneman and Tversky (1972a) suggested that a salient feature of a binomial process is the identity of the majority outcome. Thus, a binomial sample that preserves the majority-minority relation in the population is judged to be more probable than one that violates this relation, regardless of objective considerations. As an alternative hypothesis, Olson thought that perhaps people who are not knowledgeable about probabilistic sampling may treat the stated percentages as absolute numbers rather than as proportions. The sample is then judged according to whether those physical numbers could arise from the given population. For example, Olson predicts that subjects would judge that a sample with 45 % X's is more likely to come from a population with 55 % X's than from one with 35 % X's. Kahneman and Tversky would make the opposite prediction. Olson (1976) obtained support for his predictions utilizing numerous word problems.

Olson's conclusion was not that the representativeness was wrong, but rather

that it was incomplete without specifying the dimensions along which representativeness is judged. He distinguished between what he called "superficial task characteristics and underlying problem characteristics", the former depending on the wording of the question, the form of the response, etc., and the latter being that about which a judgment is sought. He concluded that his and Kahneman and Tversky's subjects were responding to the superficial task characteristics rather than to the underlying problem characteristics.

This distinction can be employed in summarizing the research on sampling distributions. First of all, the impact of sample size on judgment is frequently small, but generally not zero, and furthermore, for some people it is quite large. The primary difficulty with research in this area is the lack of specificity concerning the determinants of representativeness combined with the sensitivity of people's responses to the wording of the problems.

Discussion

The purpose of this very selected review was to indicate some of the basic effects for which a theory must account that have emerged in the study of heuristics, and some of the limitations in many of the current experimental designs. The limitations, of course, derive from insufficient attention to the effects, so the two aspects can be discussed simultaneously. The manner in which people use uncertain information to arrive to a judgment seems to depend, at least in part, on (i) how specifically the information refers to the event in question, (ii) factors that affect the salience of the information, (iii) individual differences, including expertise, training, and quantitative aptitude, and (iv) the wording of the problem. Theories are needed to elucidate these phenomena, and experimental designs are required that pay more attention to them.

Recognition of these factors is growing in the heuristics literature, as evidenced, for example, by the recent thorough discussion by Kahneman and Tversky (1982). Also, as indicated earlier, theories regarding use or interpretation of probabilistic information are being developed. However, it is difficult at this point to see any real impact of these developments on the design and interpretation of experiments on judgmental heuristics.

Furthermore, none of the theories offered thus far attempt to explain individual differences or the effects of problem wording.

As Scholz (1981) has pointed out, the difficulty in analyzing responses to specific word problems is that one cannot separate semantic considerations unique to the particular problem from the application of global probabilistic concepts. There is considerable evidence in other problem solving areas that the wording of a problem crucially determines how it is understood and the nature of the cognitive operations that will be used to solve it. Simon (1978) has suggested that a person translates a particular situation into a "problem representation" or a "problem space", which is then operated upon or about which judgments are made. The final response depends both upon the encoding of the task and the nature of the judgment process. This was demonstrated by Hayes and Simon (1977) in their study of problem isomorphs, i.e., formally identical problems that differ only in terms of the semantic encoding of the relationships. In that experiment, the nature of the semantic encoding profoundly affected subjects' solutions.

In another context, Wason and his colleagues (Wason & Johnson-Laird 1970; Wason & Shapiro 1971) have demonstrated that people have difficulty solving a simple deductive logic problem when the problem is stated abstractly, but have no difficulty at all when it is reformulated in terms of situations they are used to thinking about. Very recently, Cox and Griggs (1982) have shown that performance in the task depends on the degree of similarity of the problem content to subjects' experiences, and they propose a reasoning by analogy explanation.

There is also research within the judgment area showing that the nature of the wording influences the response, much of it summarized and discussed very thoroughly by Fischhoff, Slovic and Lichtenstein (1980). However, the implications of this work have not yet been felt in the research on heuristics.

Turning to the problem of individual differences, some attention has been given in the problem solving literature to the nature of the differences between experts and novices. For example, Chi, Feltovich and Glaser (1981) sug-

gested that when faced with a physics problem experts abstract the basic principles from the particular problem while novices represent the problem in terms of its literal features. Not surprisingly, the experts bring a good deal of tacit knowledge to the task that allows them to make inferences and derivations from the situation described by the problem statement. In a medical judgment situation, Johnson, Hassebrock, Duran and Moller (1982) demonstrated considerable differences between experienced physicians and less experienced ones in how they establish diagnoses based on available information.

The lesson behind all this is that it is not sufficient simply to ask whether a particular heuristic is used or a particular judgment is rational. The original work demonstrating heuristics was extremely important in refocussing research away from only testing normative models. However, we have now reached the point where it is necessary to develop theories of problem representation and of judgment that concern both the semantic and the statistical aspects of the situations. These theories must indicate how the problem structure determines which dimensions of the information will be encoded, and how these features will be used. The research on heuristics should rely less on individual word problems, and more on the systematic manipulation of features in a manner determined by the theory under consideration.

References

- Anderson, N.H. & Barrios, A.A.: 1961. Primacy effects in personality impression formation. In: *Journal of Abnormal and Social Psychology*, 63, 346-350
- Anderson, N.H. & Hubert, S.: 1963. Effects of concomitant verbal recall on order effects in personality impression formation. In: *Journal of Verbal Learning and Verbal Behavior*, 2, 379-391
- Asch, S.E.: 1946. Forming impressions of personality. In: *Journal of Abnormal and Social Psychology*, 41, 258-290
- Bar-Hillel, M.: 1979. The role of sample size in sample evaluation. In: *Organizational Behavior and Human Performance*, 24, 245-257
- Bar-Hillel, M.: 1980. The base-rate fallacy in probability judgments. In: *Acta Psychologica*, 44, 211-233
- Borgida, E. & Nisbett, R.: 1977. The differential impact of abstract vs. concrete information. In: *Journal of Applied Social Psychology*, 7, 258-271

- Carroll, J.S. & Siegler, R.S.: 1977. Strategies for the use of base-rate information. In: *Organizational Behavior and Human Performance*, 19, 392-402
- Chi, M.T.H./Feltovich, P.J. & Glaser, R.: 1981. Categorization and representation of physics problems by experts and novices. In: *Cognitive Science*, 5, 121-152
- Christensen-Szalanski, J.J.J. & Beach, L.R.: 1982. Experience and the base-rate fallacy. In: *Organizational Behavior and Human Performance*, 29, 270-278
- Christensen-Szalanski, J.J.J. & Bushyhead, J.B.: 1981. Physicians' use of probabilistic information in a real clinical setting. In: *Journal of Experimental Psychology: Human Perception and Performance*, 7, 928-935
- Cohen, L.J.: 1979. On the psychology of prediction: Whose is the fallacy? In: *Cognition*, 7, 385-407
- Cohen, L.J.: 1981. Can human irrationality be experimentally demonstrated? In: *The Behavioral and Brain Sciences*, 4, 317-331
- Cox, J.R. & Griggs, R.A.: 1982. The effects of experience on performance in Wason's selection task. In: *Memory and Cognition*, 10, 496-502
- Delaney, H.D. & Wallsten, T.S.: 1977. Probabilistic information processing: Effects of a biased payoff matrix on choices and bids. In: *Organizational Behavior and Human Performance*, 20, 203-237
- Einhorn, H.J. & Hogarth, R.M.: 1982a. A theory of diagnostic inference: I. Imagination and the psychophysics of evidence. Chicago, IL, Center for Decision Research, University of Chicago
- Einhorn, H.J. & Hogarth, R.M.: 1982b. A theory of diagnostic inference: II. Judging causality. Chicago, IL, Center for Decision Research, University of Chicago
- Evans, J.St.B.T. & Dusoier, A.E.: 1977. Proportionality and sample size as factors in intuitive statistical judgment. In: *Acta Psychologica*, 41, 129-137
- Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1979. Subjective sensitivity analysis. In: *Organizational Behavior and Human Performance*, 23, 339-359
- Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1980. Knowing what you want: Measuring labile values. In: Wallsten, T.S. (ed.), *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ, Erlbaum, 115-141
- Hayes, J.R. & Simon, H.A.: 1977. Psychological differences among problem isomorphs. In: Castellan, N.J., Jr./Pisoni, D.B. & Potts, G.R. (eds.), *Cognitive Theory*, Vol. 2, Hillsdale, NJ, Erlbaum, 21-41
- Irwin, F.W.: 1953. Stated expectations as a function of probability and desirability of outcomes. In: *Journal of Personality*, 21, 329-335

- Irwin, F. & Snodgrass, J.: 1966. Effects of independent and dependent outcome values upon bets. In: *Journal of Experimental Psychology*, 71, 282-285
- Jepson, C./Krantz, D.H. & Nisbett, R.E.: 1982. Inductive reasoning: Competence or skill? (Bell Labs Technical Memorandum). Murray Hill, NJ, Bell Labs
- Johnson, P.E./Hassebrock, F./Duran, A.S. & Moller, J.H.: 1982. Multimethod study of clinical judgment. In: *Organizational Behavior and Human Performance*, 30, 201-230
- Kahneman, D. & Tversky, A.: 1972a. Subjective probability: A judgment of representativeness. In: *Cognitive Psychology*, 3, 430-454
- Kahneman, D. & Tversky, A.: 1972b. On prediction and judgment. (ORI Research Monograph, 12 (4)) Eugene, OR, Oregon Research Institute
- Kahneman, D. & Tversky, A.: 1973. On the psychology of prediction. In: *Psychological Review*, 80, 237-251
- Kahneman, D. & Tversky, A.: 1982. On the study of statistical intuitions. In: *Cognition*, 11, 123-141
- Krantz, D.H./Fong, G.T. & Nisbett, R.E.: 1982. Formal training improves the application of statistical heuristics to every day problems (Bell Labs Technical Memorandum). Murray Hill, NJ, Bell Labs
- Manis, M./Dovalina, I./Avis, N.E. & Cardoze, S.: 1980. Base rates can affect individual predictions. In: *Journal of Personality and Social Psychology*, 38, 231-248
- Mynatt, C.R./Doherty, M.E. & Tweney, R.D.: 1977. Confirmation bias in a simulated research environment: An experimental study of scientific inference. In: *Quarterly Journal of Experimental Psychology*, 29, 85-95
- Nisbett, R.E./Borgida, E./Crandall, R. & Reed, H.: 1976. Popular induction: Information is not necessarily informative. In: Carroll, J.S. & Payne, J.W. (eds.), *Cognition and Social Behavior*, Hillsdale, NJ, Erlbaum, 113-134
- Olson, C.L.: 1976. Some apparent violations of the representativeness heuristic in human judgment. In: *Journal of Experimental Psychology: Human Perception and Performance*, 2, 599-608
- Scholz, R.W.: 1981. Stochastische Problemaufgaben: Analysen aus didaktischer und psychologischer Perspektive. IDM - Materialien und Studien Band 23. Bielefeld, Universität Bielefeld
- Shafer, G.: 1981. Constructive probability. In: *Synthese*, 48, 1-60
- Simon, H.A.: 1978. Information-processing theory of human problem solving. In: Estes, W.K. (ed.), *Handbook of Learning and Cognitive Processes* Vol. 5: Human Information Processing. Hillsdale, NJ, Erlbaum, 271-296

- Tversky, A. & Kahneman, D.: 1971. The belief in the law of small numbers. In: *Psychological Bulletin*, 76, 105-110
- Tversky, A. & Kahneman, D.: 1973. Availability: A heuristic for judging frequency and probability. In: *Cognitive Psychology*, 5, 207-232
- Tversky, A. & Kahneman, D.: 1974. Judgment under uncertainty: Heuristics and biases. In: *Science*, 185, 1124-1131
- Tversky, A. & Kahneman, D.: 1980. Causal schemas in judgments under uncertainty. In: Fishbein, M. (ed.), *Progress in Social Psychology*, Vol. 1. Hillsdale, NJ, Erlbaum, 49-72
- Wallsten, T.S.: 1976. Using conjoint measurement models to investigate a theory about probabilistic information processing. In: *Journal of Mathematical Psychology*, 14, 144-185
- Wallsten, T.S.: 1977. Processing information for decisions. In: Castellan, N.J./Pisoni, D.B. & Potts, G. (eds.), *Cognitive Theory*, Vol. II. Hillsdale, NJ, Erlbaum, 87-116
- Wallsten, T.S.: 1981. Physician and medical student bias in evaluating diagnostic information. In: *Medical Decision Making*, 1, 145-164
- Wallsten, T.S. & Barton, C.: 1982. Processing probabilistic multidimensional information for decisions. In: *Journal of Experimental Psychology: Learning Memory and Cognition*, 8, 361-384
- Wason, P.C. & Johnson-Laird, P.N.: 1970. A conflict between selecting and evaluating information in an inferential task. In: *British Journal of Psychology*, 61, 509-515
- Wason, P.C. & Shapiro, D.: 1971. Natural and contrived experience in a reasoning problem. In: *Quarterly Journal of Experimental Psychology*, 23, 63-71

This Page Intentionally Left Blank

THE BASE RATE FALLACY CONTROVERSY

MAYA BAR-HILLEL

*Hebrew University
Department of Psychology
Israel*

It is not easy to say something new and enlightening about the base-rate fallacy controversy, since only recently there have appeared various summaries and reviews of it (e.g., Borgida & Brekke 1981; Kassin 1979; Tversky & Kahneman 1982a). Moreover, I doubt that there really is much of a controversy, as such, surrounding the issue. In what follows, I will attempt to sketch a conceptual and experimental history of the base rate issue, hoping to show that the emergent picture is really remarkably coherent.

The base rate fallacy and attributions of causality

To begin, I shall give three puzzles that illustrate the base rate fallacy, but in a context of causal attribution rather than probability estimates, these two often being intimately linked to each other. The task is to explain each of the following facts.

1. A children's riddle asks: Why is more grass consumed by white sheep than by black sheep?
2. A Rhode Island newspaper proclaimed that in 1957, 10 % of all pedestrians killed had crossed an intersection on a green light, while only 6 % had crossed on a red light (Huff 1959).
3. In 1973, the University of California at Berkeley accepted 44 % of all

male applicants, versus only 35 % of all female applicants. Yet the females were overall the more qualified applicants, and there was no sex discrimination practiced (Freedman, Pisani & Purves 1978).

Most people are inclined to look for causal explanations of the above facts, which leads to a startling conclusion in the second example, and bewilderment in the third. In fact, however, all three can be explained by reference to base rates (hint: there may just be more white than black sheep).

The base rate fallacy in the evaluation of diagnostic test results

The importance of considering base rates before making causal attributions is one which is instilled in us all in the course of our training in experimental methodology. But base rates play an important role in other inferential formats, too, especially Bayesian ones. Yet there is evidence aplenty that in those contexts they are largely ignored in favor of the diagnostic information at hand. This is the phenomenon known as the base rate fallacy.

As late as in 1955, Meehl and Rosen found it necessary to caution against committing the base rate fallacy in the context of evaluating a diagnostic-test result. Even a fairly accurate diagnostic device, they pointed out, could lead to conclusions that are less often right than wrong (and certainly less often right than judging solely by base rates), unless its results are properly integrated with other factors, most notably base rate considerations. Yet according to Dershowitz (1971), Eddy (1982), Foltz and Kelsey (1978) and Lykken (1975), to cite but a few, this fallacy is still prevalent in such important areas of judgment and decision making as preventive detention, mammography, Pap smears, and lie detection, respectively.

The first laboratory studies

For a psychologist, a natural question to ask is "why?". Why are base rates so easy to disregard in such settings? Kahneman and Tversky, founders of the heuristics approach to judgment under uncertainty, were the first to study the base rate fallacy in a controlled laboratory setting, and to attempt a cognitive account of it. Interestingly, however, the link with the above

types of problems was only made later, since Kahneman and Tversky's interest in the base rate fallacy originated from a prediction generated by their representativeness hypothesis (Kahneman and Tversky 1972a), rather than from an attempt to account for its above-mentioned manifestations.

According to this hypothesis, the probability of uncertain outcomes is often judged by the extent to which they represent their source or generating process. The features by which this similarity or representativeness is assessed are not necessarily those that figure in the normative derivation of the requested probability. From this hypothesis, Kahneman and Tversky derived the following prediction: If an event is to be judged vis a vis several alternative possible sources, or several alternative possible outcomes, these will be ranked by the similarity between them and the event (or vice versa), and the ranking will not be affected (sufficiently, or at all) by how likely each source or outcome is *initially*.

This prediction was borne out by a series of studies, one of the best known of which is the Tom W. study (Kahneman & Tversky 1973). One Tom W., known only by a brief and admittedly unreliable personality sketch, was judged to be more likely to be in Computer Science than in Social Sciences by subjects who were well aware how much more common the latter kind of student is, merely because the sketch cleverly evoked a computer-buff stereotype more than a social-worker stereotype. Although the normative analysis of the Tom W. task proceeds via Bayes' Theorem, as it does in the contexts discussed by Meehl and Rosen and the others, no link was established between them in the 1973 paper, and the expression "base rate fallacy" was not used.

Actually, though, by that time Kahneman and Tversky already had the first results of the later notorious Cab Problem (Kahneman & Tversky 1972b). This problem also shows how diagnostic evidence (an eyewitness' testimony rather than a psychologist's personality sketch) dominates base rate considerations. It was not included in the 1973 paper, though, since it really was not a case of judgment by representativeness.

Although the 1973 results have stood up remarkably well to the test of time, replication, and even refutation efforts, two points made by Kahneman and

Tversky had to be modified later. One was the implication that base rates were ignored *because* they do not affect representativeness. The second was based on results obtained in their Engineer/Lawyer study. In this study, subjects guessed the profession of five people on the basis of brief personality sketches. One sketch was completely nondiagnostic with respect to these two professions. Subjects gave a 50 % - 50 % chance that its owner, Dick, was an Engineer, rather than the probability indicated by the categories' base rate. From this, Kahneman and Tversky concluded that even when representativeness fails to be a guide to prediction (e.g., even when the judged event is equally representative of all contending outcomes), it dominates base rate considerations.

That the base rate fallacy is not necessarily caused by the representativeness heuristic is shown by the results of the Cab Problem, as well as in the classical diagnostic-test-result-on-actuarial-background cases, where the necessary probabilistic components are given, and need not be assessed by representativeness. The second point was challenged by Ginosar and Trope (1980). In part, their study was a direct replication of Kahneman and Tversky's Engineer/Lawyer study, but Ginosar and Trope's subjects each judged only one of the five sketches. In contrast to Kahneman and Tversky's results, Ginosar and Trope found that the nondiagnostic sketch *was* assessed by the base rate. On the basis of their results, they concluded that, when the individuating information is useless as a guide to prediction, "attention is shifted ... to the base rate frequencies" (p. 240). It seems that Tversky and Kahneman's subjects judged Dick equally likely to be an Engineer or a Lawyer under both base rate conditions only because they were given a multiple prediction task; the usefulness of representativeness to the evaluation of most sketches created some sort of "set".

Tversky and Kahneman have accepted both these modifications. They concede that "the neglect of base rates appears to be a more general phenomenon, which occurs even when probability assessments are not mediated by similarity" (Tversky & Kahneman 1980, p. 79), and that "Most the available data ... support a more moderate hypothesis that intuitive predictions and probability judgments are highly sensitive to representativeness, although they are not completely dominated by it" (Tversky & Kahneman 1982b, p. 88).

So the first round in the base rate fallacy controversy ends with the conclusion that while representativeness considerations are often overweighted and base rate considerations are often underweighted, the issue of when, why and to what extent base rates are neglected demands a logically separate account (Bar-Hillel 1981). The results strongly underscored, however, that people at no time actually predict against a strong representativeness indication.

Two experimental paradigms

I shall call the study of problems in which the base rate fallacy is divorced from representativeness "the textbook-problem paradigm", while the paradigm established by Kahneman and Tversky will be called "the social-judgment paradigm". Let us now follow the development of these two paradigms separately, noting parallels as they emerge.

Ignoring base rates versus inverting conditional probabilities

Before attempting a theory of base rate usage, it is worth noting that often judgments may have the appearance of manifesting the base rate fallacy and yet result from a cognitive dynamic that has nothing to do with the ignoring of base rates. According to Bayes' Theorem, which is the normative formalization of most of the problems developed by both paradigms, the posterior probability of an event, $p(E/D)$, depends both on data impact, $p(D/E):p(D)$, and on the event's base-rate, $p(E)$. On occasion, however, $p(E/D)$ can be estimated directly, rather than via Bayes' Theorem. In the Cab Problem, for example, subjects were asked for the probability that the errant cab was green following the witness' testimony to that effect, $p(G/"G")$. This conditional probability is easily confused with its inverse, $p("G"/G)$ (perhaps since both are natural formalizations of "the probability of an accurate Green cab identification"). Hence, it is quite possible that some of the Cab Problem subjects were not ignoring base rates, but merely confusing two conditional probabilities. This is an error, and it leads to the same final response, but is not the error of ignoring base rates.

A parallel possibility emerges in the Tom W. paradigm. Recently, I gave subjects the Tom W. task, but requested them to rank the probabilities that

students in different fields of study would be describable like Tom W., $p(D/F)$, rather than requesting, as Kahneman and Tversky did, a ranking of the probabilities that someone describable like Tom W. is enrolled in the different fields of study, $p(F/D)$ (Bar-Hillel 1980). Ten years after the original study, my subjects' rankings correlated .98 with Kahneman and Tversky's! The point I want to make is that it is difficult to say whether Kahneman and Tversky's subjects were judging $p(F/D)$ (by representativeness) directly - in which case base rates are really irrelevant, or were judging $p(D/F)$ (also by representativeness), and then failing to integrate it with base rate considerations.

The confusion between a conditional probability and its inverse is not limited to the psychological laboratory. In a startling paper, Eddy (1982) documents that the medical literature dealing with mammograms is full of just such confusions (e.g., between p (disease/positive test-result), called predictive accuracy, and p (positive test-result/disease), called retrospective accuracy). In other words, some of the judgmental errors in mammogram interpretation may be the result of the confusion of two different conditional probabilities, rather than the ignoring of base rates.

The potential for such confusion has somehow been ignored in the published discussions of the base rate fallacy (Bar-Hillel 1980). Perhaps universal explanations were sought, and it is so obvious that this is not a universally occurring confusion. No one, presumably, would fail to distinguish between the probability of being male if one is a head of state (which is close to one), and the probability of being a head of state if one is male (which is close to zero). Not even all empirical results where probability judgments show insensitivity to base rates can be accounted for by such a confusion. In the Suicide Problem (Bar-Hillel 1980), for example, subjects were informed that among the causes of death in young adults, p (suicide/married) = 5 %, while p (suicide/single) = 15 %. They estimated p (married/suicide) to be 25 %. This response ignores the base-rate of married people in the relevant population, but at the same time, it differs numerically from the value of the inverse conditional probability. Similarly, "McCauley and Stitt (1978) gave subjects a list of nine traits (e.g., "efficient") and asked them to estimate both p (German/Trait) and p (Trait/German) ... They found a product

moment correlation of .83 between the two ... At the same time, mean $p(\text{Trait}/\text{German})$ was between 1.7 and 3.1 times higher than mean $p(\text{German}/\text{Trait})$ " (Bar-Hillel 1981, p. 7).

The causality factor

So perhaps rather than asking why or when base rates will be ignored - which presupposes that they are - it is better to ask when they will be incorporated. The first factor that was identified as enhancing the impact of base rates was causality.

Within the social-judgment paradigm, the effect of causal base rates was shown by Ajzen (1977). A student, Gary W., was described by a brief description. Subjects were asked to estimate Gary's probability of having passed some test. Some subjects were told that Gary had been randomly sampled from a group of students amongst whom it just so happened that 75 % had failed the test and 25 % had passed. These subjects ignored the base rate of passing, even though it is normatively relevant. Other subjects were told that this test was, overall, failed by 75 % of students and passed by 25 % of students in Gary's class. These subjects were responsive to the base rate of test failures, presumably because here the base-rate could be causally tied to the target event readily, as in the inference that the test was a difficult one.

A parallel phenomenon occurs in the textbook-problem paradigm. Typically, the Cab Problem states that 85 % of the cabs in some town are Blue. This information is ignored. But subjects who are told that both cab companies are of equal size, and yet the Blue ones account for 85 % of the accidents in that town, no longer ignore the base rate information (Tversky & Kahneman 1980).

The causality hypothesis can be strengthened, from merely predicting that causal base rates affect judgments more than noncausal ones, to predicting that in problems where both are presented, causal base rates will dominate noncausal ones, even if the latter are objectively diagnostic. This, indeed, is just what happened in the Suicide Problem described earlier. Note that the weak version of the hypothesis predicts the relative magnitude of the

base rate effect between problems, whereas the strong version makes a prediction within a problem. There cannot be a test of this in the social-judgment paradigm, since there, by definition, one of the items of information is not a base rate.

The specificity factor

The second hypothesis was initially formulated as a within-problem one. It states that more specific information will tend to dominate a less specific one. Note that in the social-judgment paradigm, the individuating information not only lends itself to judgment by representativeness, but in addition pertains to the target case more specifically than the base rate - so specifically, indeed, as to be individuating. Perhaps this is why the output of the representativeness judgment is not then integrated with the base rate: being more specific, it just dominates it.

There are independent reasons for believing such a hypothesis. In many cases, it is normative to replace a general base rate by a more specific one. For example, for the purpose of predicting the life expectancy of a newborn German baby, the life expectancy in Germany replaces the worldwide life expectancy. Another example is given in the Modified Cab problem (Bar-Hillel 1980). In this problem, subjects were given, in addition to the city base rates and in place of the witness, the base rates of Blue and Green cabs in the vicinity of the accident. The pattern of responses given to this problem was much like that given to the standard problem, that is, the median subject ignored the city base rate in favor of the more specific information - only in the Modified Cab problem, this is the appropriate thing to do.

It seems likely that people, especially those untutored in statistics and probability, would find it hard to distinguish between Bayesian problems, where the diagnostic information - even if it is more specific than the base rate - does not replace it, and between problems such as the Modified Cab problem. This might be the origin of the specificity factor ¹⁾.

¹⁾ This factor can also account for the results of problems such as the Dreamers Problem and the Urn and Beads Problem (Bar-Hillel 1980), which can not be accounted for by any of the other explanations (representativeness, causality, confusion of conditional probabilities, etc.).

Causality and specificity considered jointly

Combining the causality and the specificity factor yields a testable implication: If two base rates are presented in a (Bayesian) problem and no one of them is either more specific or more causal than the other, then neither will dominate, and the two will be integrated. Evidence supporting this prediction is to be found in the Intercom Problem (Bar-Hillel 1980). This is a cab problem in which the errant cab is identified as one that has an intercom in it, and the two cab companies are known to have a different proportion of cabs equipped with intercoms. This differential base rate can no more be causally tied to accident proneness than company size, nor does it pertain to a more specific subset of the overall population. And, lo!, it does not dominate the overall city base rate.

The vividness factor

Meanwhile, another factor had been proposed to account for the base rate fallacy, which is only applicable to the social-judgment paradigm. This is the factor of vividness or saliency. Nisbett and Borgida (1975) argued that a base rate is "abstract and pallid", while individuating information is "concrete and vivid" and therefore more memorable and utilizable. Like the previous two, this factor can also generate two predictions: i. By making base rates more vivid their effect can be increased. ii. Base rates that are as vivid as the individuating information will not be dominated by it.

A classical example of the effect of vividness is the Volvo effect (Nisbett, Borgida, Crandall & Reed 1976, p. 129): A person who has decided to buy a Volvo on the basis of the statistics gathered by Consumers Reports, changes his mind upon overhearing, in a cocktail party, the woeful, but lively, description of the demise of one Volvo purchased by the teller's brother-in-law.

In a penetrating look at the vividness factor, Taylor and Thompson (1982) concluded that, although "case history information does have a persuasive impact on judgments when it is contrasted with base-rate or abstract information ... the extent to which this finding speaks to the vividness effect is

questionable ... (p. 162).

One study provided direct evidence supporting Taylor and Thompson's conclusion. Manis, Dovalina, Avis and Cardoze (1980) tested whether in problems where it has been established that people consider base rates to be irrelevant, their consideration can be altered by making the base rates more vivid. They used the social-judgment paradigm, presenting base rates in the standard numerical summary form to some subjects, but via a step-by-step labelled presentation of the population to others. The latter base rates were not more influential than the former, in contrast to the vividness prediction.

Christensen-Szalanski and Beach (1982) gave another twist to the vividness hypothesis. They conjectured that to increase subjects' sensitivity to base rates, not only the base rates themselves must be vividly experienced, but also their relationship to the diagnostic information. To that end, their subjects were introduced, slide by slide, to a population of diseased or healthy people, who had obtained positive or negative results on a diagnostic test for detecting that disease. Subjects were ultimately asked to assess $p(\text{disease/positive test-result})$. Assessments were quite reasonable, which the investigators took to be evidence for their conjecture. But as Beyth-Marom and Arkes (1983) correctly pointed out, it is eminently possible that subjects were simply directly assessing the required conditional probability from the data base they had been exposed to. Beach himself (Peterson & Beach 1968) had demonstrated that people are quite proficient at extracting summary statistics from data. If so, they had no need for Bayesian reasoning in Christensen-Szalanski's and Beach's task, hence the observed improvement in their performance has no bearing on the conjecture that direct experience aids in overcoming the base rate fallacy.

Multiple judgments tasks

Up to now, all the factors we have explored pertain to a single judgment. When subjects are called upon to make a series of judgments, other factors enter the picture. The most important is that subjects may want the distribution of their multiple judgments to reflect the population base rate. This is a variant of representativeness, but one which focusses on base rates

rather than ignoring them. Another possibility is that subjects enter a response set, namely that they adopt a certain fixed convenient strategy throughout the judged stimuli, though they would have adopted different strategies if they were making a single judgment. This possibility was earlier raised when explaining Kahneman and Tversky's "Dick" results. Lastly, when considering multiple judgments tasks, it is important to remember that their results are pooled across both subjects and stimuli, a fact that might influence the interpretation.

With these possibilities in mind, let us review the multiple judgments studies. Such studies have only been done within the social-judgment paradigm.

The probability matching paradigm

The first factor mentioned above, that subjects making multiple judgments may feel called upon to distribute their judgments in a manner that represents the population base rate, is known as "probability matching". In probability-matching experiments, subjects sequentially predict the identity of each item in a binomial series. Typically, they learn the base rate of the item categories from their trial-by-trial feedback, and eventually come to match it in their predictions (see, e.g., Estes 1976). In the probability-matching experiments, however, there was nothing to distinguish one trial from the next - except the pattern of preceeding trials. In contrast, judgment studies that call for sequential predictions give an item of individuating information on each trial. Thus, the probability-matching paradigm and the social-judgment paradigm elicit different response tendencies - matching base rates and judging by representativeness, respectively.

Representing stereotypes versus representing base rates - a conjecture

These two tendencies, however, need not be incompatible. If the body of judged stimuli contains a mixture of representative and non representative stimuli, a subject can classify the first according to representativeness, and simultaneously probability-match by classifying the neutral stimuli in a manner that will give an overall correspondence between prediction proportions and population proportions. And what if the mixture does not allow

both tendencies to be satisfied at once? This has not been tested yet, but Bar-Hillel and Fischhoff (1981) ventured the hypothesis that subjects will not violate representativeness unless that leads to a sample that is impossible, rather than merely improbable, given the population.

Indirect evidence for the conjecture

What evidence do we have to support the conjecture that subjects strive simultaneously to judge by representativeness, and to probability match? A study by Carroll and Siegler (1977) required subjects to classify 20 short personality descriptions as belonging to either an engineer or a lawyer. The way the descriptions were constructed makes it unlikely that any would strongly resemble either stereotype. When subjects were told that, among the 20 descriptions, 70 % were members of one profession (engineers, for one group of subjects, and lawyers, for another) and 30 % belonged to the other, an average of 62 % of the classifications were into the majority profession. When told that the 20 were sampled at random from 400 descriptions which were divided between the two professions in a 70 %/30 % ratio, only 56 % of the classifications, on average, were into the majority profession. Lastly, when ten descriptions were designed to resemble, respectively, five lawyers and five engineers, they were all classified into the category they were intended to resemble by an average 65 % of the subjects (70 % when the category was a majority one, but also 60 % when it was a minority one).

Since the extent to which descriptions resembled either category stereotype was not independently ascertained in Carroll and Siegler's study, it is conceivable that their data resulted from having different subjects disagree which profession a given description represents, if indeed any, but having no subject classify descriptions contrary to representativeness. (Note that when results are pooled over both subjects and stimuli, what might seem "weakly representative" stimuli may simply be stimuli that are judged as representative by fewer subjects than "strongly representative" stimuli. But I have deliberately refrained from distinguishing between levels of representativeness, and try, as far as possible, to explain all results by the dichotomy representative/not representative).

Although Carroll and Siegler found a consistent base rate effect in their studies, the effect falls short of probability matching. This could be due to several reasons. a. The subjects were dealing with a small number of predictions relative to that customarily rendered in probability matching experiments. Probability matching emerges gradually rather than immediately in those studies. b. The sample judged in Carroll and Siegler's study was a small physical subset from some fixed population, rather than being the step-by-step realization of a random generating-process. In probability matching experiments, subjects aim for predictions that will be "just like" the actual sequence. Possibly in classifying samples, especially such small ones, their only concern is not to produce a sample distribution that is too incompatible with the population, as we know that subjects believe random samples should represent sampling error as well as population characteristics (Kahneman & Tversky 1972a).

Manis et al. (1980) followed the probability-matching paradigm more closely, by forcing subjects to learn the base rates trial by trial, in a long series (50) of predictions. Indeed, they found a larger base rate effect. In the last block of ten trials, 70 % of their subjects' predictions were into the category enjoying the 80 % majority.

Manis et al. contrasted their result with what they took to have been Kahneman and Tversky's representativeness-based prediction. In fact, however, Bar-Hillel and Fischhoff (1981) showed that Manis et al.'s results are perfectly compatible with previous results, according to the conjecture put forth above, namely, that every subject classified those stimuli that were clearly representative of some category into that category, irrespective of the base rate, whereas nonrepresentative stimuli were classified largely into the majority category, resulting in a proportion of majority category predictions that all but match the category proportion. Indeed, working from estimates given by Manis et al. concerning the proportion of neutral versus representative stimuli in their study, Bar-Hillel and Fischhoff used the conjecture to predict the proportions of various classifications to be expected under the various base rate conditions, and came remarkably close to the proportions actually observed (see Figure 1).

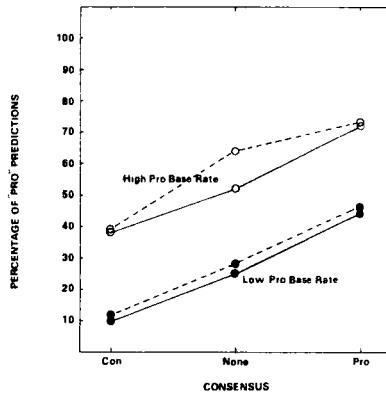


Figure 1: Predictions (solid line = actual, broken line = reconstructed) concerning category membership of stimuli characterized by individuating information as a function of the observed base rate and the consensual classification of the stimuli by representativeness. (From "When do base rates affect predictions?" by M. Bar-Hillel and B. Fischhoff, *Journal of Personality and Social Psychology* 41 (1981) 671-680, Figure 1).

Earlier I mentioned Manis et al.'s study as one which failed to produce a greater base-rate effect for more vivid base rates. Indeed, we now see that their entire effect can be attributed to probability matching tendencies, while vividness contributes nothing.

Direct evidence for the conjecture

We have seen that the data of previous multiple-judgments studies are compatible with our conjecture. The time had come for testing it directly. To that end, subjects were given ten short personality descriptions (Fischhoff & Bar-Hillel 1981). They assessed how business-executive like versus university-professor like the descriptions really were. Over 75 % of the subjects agreed on five of the descriptions (four of them executives, one professor), but there was considerable difference of opinion concerning the five others. Surprisingly, however, even those descriptions that were judged executive-like by some subjects and professor-like by others, were judged so with considerable confidence, and were rarely judged as neutral (only one descrip-

tion was judged as neutral by over 10 % of the subjects). Even though a test of our conjecture requires a comparison of how representative versus neutral stimuli are judged, and our stimuli turned out to be not neutral enough, the data nevertheless supported the conjecture, as follows.

In the course of making the prediction categories increasingly similar (from executive/professor through engineer/physicist to dentist/orthodontist), the percentage of neutral judgments rose from 8.4 % through 24.0 % to 31.5 %, respectively. Correspondingly, the base rate effect increased from 4.5 % (which is much like the magnitude of effect reported by Kahneman and Tversky, 1973, for their engineer-lawyer categories) through 8.7 % to 12.0 %, respectively. If subjects were probability matching on the neutral stimuli, and judging by representativeness elsewhere, the size of the effect would have gone from 3.4 % through 9.6 % to 12.6 %, respectively. These figures are much like the ones actually observed. We are presently engaged in an attempt to put the conjecture to a stronger test.

The base rate fallacy and procedural variables

In their recent review of the base rate issue, Tversky and Kahneman (1982a) distinguish between evidential variables - those that "refer to the nature of the source and the interpretation of the evidence" (p. 154), such as the relative relevance of the given base rates, and procedural variables - those that "refer to the properties of the design, the task, and the display" (p. 154), while leaving the actual information given unchanged. Hitherto, we have dwelt on evidential variables. Many different procedural variables have also been studied. Overall, it is fair to generalize that they have had very disappointing effects, that is, none at all, or such that were spurious and fleeting, neither resulting from nor resulting in greater understanding of the normative role of base rates. Let us observe one example in detail.

A study by Fischhoff, Slovic and Lichtenstein(1979) attempted to increase the saliency of the base rate in the Cab Problem by exhorting subjects to perform a sort of "subjective sensitivity analysis", called SSA. Subjects were asked to consider how different base rate values might affect their summary judgments. Subjects exhibiting the base rate fallacy in the extreme

would have replied consistently with the witness' accuracy, 80 %. Instead, mean responses were monotonically related to the shifting base rates, and though they were, appropriately, lower than 80 %, they were not nearly low enough. Moreover, the base rate fallacy was reinstated as soon as these same subjects were given a formally equivalent, but differently worded, problem, such as Lyon and Slovic's (1976) Light Bulb problem.

To check whether this poor showing was nevertheless an improvement, Fischhoff and Bar-Hillel (1982) had subjects do an SSA on the Modified Cab problem, in which ignoring the city base rate is the right thing to do, and on the Dreamer problem, in which the right thing to do is to use the Dreamers base rate only. Subjects repeated the pattern that had been previously observed. They gave mean responses that were lower than 80 %, varied monotonically with the base rate - and were totally unrelated to the normative strategy. "It was as if the variable manipulated in a SSA only succeeded in saying to subjects 'use me', rather than 'comprehend me'" (Fischhoff & Bar-Hillel 1982, p. 16). Two other "Focussing Techniques", similar to SSA, were tried out with similarly dismal results.

Other variables that have yielded, at best, mixed and unreliable effects, if any, are: order of information presentation; translatability of the base rate in the sample; sequential rather than summary presentation of the base rates; base rate values that are compatible with knowledge or private theory versus those that aren't; and more.

The normative controversy

It is doubtful whether all of the above really describes a controversy. Rather, it is an expanding of knowledge, a deepening of understanding, an exploration of boundaries. Here and there, of course, there have been, and still exist, genuine disagreements, but typically of the kind that can be settled by reference to empirical findings. There does exist, however, a more general controversy, concerning the legitimacy of the entire heuristics research program. One of the most vocal critics of this approach is the Oxford philosopher L.J. Cohen. The base rate fallacy is but one component in this controversy, but it will serve nicely to demonstrate its essence.

Cohen classifies the biases and fallacies found in judgment and reasoning, whether probabilistic or logical, into four categories, only the first two of which acknowledge the existence of any error. The base rate fallacy enters under "Applications of inappropriate normative theory", a category that absolves the subjects of error, and places it instead on the experimenters shoulders (as does an explicit title of Cohen's: "On the psychology of prediction: Whose is the fallacy?", 1979). Specifically, Cohen goes through the Bayesian reasoning behind the Cab Problem, and then criticizes it thusly: "The ratio 17/29 is the value of the conditional probability that a cab-color identification by the witness is incorrect, on the condition that it is an identification as green. Jurors, however, ... ought not to rely on that probability ... [but rather on] the probability that a statement to the effect that the cab actually involved in the accident was green, is false, on the condition that the statement is made by the witness. If the jurors know that only 20 % of the witness's statements about cab colors are false, they rightly estimate [$p(G/"G")$] as 1/5 ... The [base rate] is strictly irrelevant. ... A probability that ... is based on causal properties, such as the physiology of vision, cannot be altered by ... chance distributions, that have no causal efficacy to the individual events" (Cohen 1981, p. 328, 329).

Cohen goes on to suggest that "the literature under criticism is propagating an analysis that could increase the number of deaths from a rare disease" (p. 329). For suppose you are suffering from one of two diseases, A and B, the base rates of which are 85 % versus 15 %. Suppose further that there exists a diagnostic test that identifies each disease correctly 80 % of the time but errs 20 % of the time. The test reports you to be suffering from B, whose base rate is only 15 %. "Should you", asks Cohen menacingly, "... opt for the treatment appropriate to A, on the supposition [reached by calculating as the experimenters did] that the probability of your suffering from A is 19/23? Or should you opt for the treatment appropriate to B, on the supposition [reached by calculating as the subjects did] that the probability of your suffering from B is 4/5? It is the former that would be the irrational one for you ... not the latter" (p. 329).

Cohen's argument does not end here, but the fundamentals must by now be clear. The controversy arises because of a basic renouncing of Bayesian reasoning,

and an upholding of those factors which we regarded as mere descriptive explanations of the base rate fallacy as normatively sound. Indeed, Cohen offers an alternative, "Baconian", probability calculus to replace the familiar, "Pascalian", one as a normative theory of subjective belief, with the additional bonus that it is more in line with prevalent human intuitions, and in particular, condones the response patterns that we have termed "the base rate fallacy".

Cohen attaches a great deal of significance to the manner in which various probabilistic arguments are regarded in court. The Law of Evidence, as he sees it, represents the distilled and reflective wisdom of many people over many generations, and cannot be shrugged off by psychologists as being mere errors of reasoning. Therefore, a probabilistic procedure upheld by the law must, by definition, be normatively correct. And the law of evidence does indeed seem to "discriminate" against base rates. Sometimes they are ruled inadmissible; other times they are given less weight than comparably, or even less, diagnostic evidence from eyewitness testimony; and they are seldom allowed to exclusively settle a case, though they may seem probative enough to do so.

This does not necessarily grant Cohen's point, however. A careful examination of the rationale for the court's attitudes to base rate information would take us too far afield, although I hasten to say that I do not share Cohen's confidence in the inherent wisdom of all laws. Nevertheless, a couple of examples are in order. "Imagine a rodeo into which 400 people are known to have been admitted through an automatic turnstile after paying the proper sum. Then 1000 people are counted on the seats, and a hole is discovered in the fence" (Cohen 1979, p. 315). It would be highly unjust, says Cohen, if indeed possible, for a jury to convict a randomly selected person in the rodeo audience on the basis of this evidence alone, though it seems that the balance of the probabilities favors the suing management. Rather than deal with this hypothetical example, consider the following example, based on a real case (Smith vs. Rapid Transit, Inc. 1945). Plaintiff was negligently run over by a blue bus, and could prove that the defendant operates four fifths of all the blue buses in town. Plaintiff could not testify to the actual identity of the bus. "That datum ... suggests that, in the absence

of other information, ... there is a "four-fifths certainty" that the defendant's bus hit this Plaintiff" (Tribe 1971, p. 1346). The absence of other information is rare, and, especially in a court of law, the very absence of any other evidence "may properly be treated as dispositive in some situations" (ibid., footnote 54). But even if the absence of any other evidence, on either litigant's part, were to be satisfactorily explained, and in a manner that leaves the probability of 4/5 unchanged, the ruling is "entirely sensible if understood ... as insisting on the presentation of some "individualized" proof of identity before compelling a party to pay damages" rather than insisting on "the minimization of the total number of judicial errors in situations of this kind" (ibid., p. 1341, footnote 37). And clearly, the law of evidence has goals that extend way beyond this last.

The normative role of base rates

Tversky and Kahneman (1982a) concluded their discussion of the evidential impact of base rates thus: "The major conclusion of [the base rate] research is that the use or neglect of consensus information in individual prediction depends critically on the interpretation of that information" (p. 160). My own conclusion is in complete agreement with theirs. Moreover, I would like to close with an example that shows, with much elegance, how the normative role of consensus information (i.e., base rates) also depends critically on its interpretation. This example as well as much of my knowledge and views on the base rate issue, is due to Daniel Kahneman and Amos Tversky.

A young doctor is moonlighting by working on holiday weekends in the emergency ward of a large urban hospital. A patient comes in, complaining of various symptoms. These symptoms look to the doctor like those signalling the onset of heart trouble, but the doctor is aware that a mere stomach disorder may sometimes exhibit the same symptoms. Experience has taught the doctor that the relative base rate of stomach disorders to heart trouble among patients is considerably higher during ordinary weekdays than during holiday weekends. How should the doctor evaluate the chances that the patient is suffering from heart trouble rather than a stomach disorder?

Intuitively, it might seem that identical symptoms should be identically eva-

luated, with no need to first check the calendar to see what the date is. But the doctor versed in the literature on judgment under uncertainty might consider it fallacious to ignore the difference in base rates, and conclude that the exact same symptoms should be judged more likely to result from heart trouble on a weekend than on a weekday. Is this indeed necessary?

The surprising answer is that both procedures may be justifiable, depending on the causal structure that leads to the different base rates on weekends and weekdays. According to one possible account, hospital emergency wards get more than the regular share of false alarms (i.e., mild complaints such as stomach disorders parading as heart trouble) on weekends, because of the relative nonaccessability of everyday medical "filters" such as neighborhood clinics, family doctors, etc. A different account is that eating habits on holiday weekends are different enough from everyday habits to genuinely increase the overall occurrence of stomach disorders. If one believes the former account, one may as well use the weekday statistics, *if a patient with the symptoms at hand* would have been sent to a hospital on a weekday anyway (making it irrelevant that others would be diagnosed as suffering from stomach disorders and never sent on to a hospital). If, on the other hand, one believes the latter account, then the population of weekend diseases is truly distributed differently than the weekday population, and doctors would be wise to take that into consideration (as they would take, say, the patient's sex into consideration, if the two sexes have different rates of heart trouble).

The indeterminacy of normativeness?

It would seem, on the basis of the case of the moonlighting doctor, that any use of base rates could be justified, provided one can come up with the proper scenario - a somewhat baffling conclusion. Happily, however, we are not faced with complete normative indeterminacy. "Proper scenarios", if by "proper" we mean "plausible", too, may not be easy to come by (see, e.g., Fischhoff, Slovic & Lichtenstein 1978). Indeed, among people with a certain shared cultural and intellectual background, a consensus typically exists concerning what is or isn't a proper scenario, therefore what is or isn't an acceptable solution to some problem. The example does show that we may need to relax our concept of a "normative solution". Sometimes a problem may have

no unique normative solution, sometimes what consensually appears to be the best solution may differ from the formal, "textbook", solution, etc. Nevertheless, some solutions are unacceptable by any sensible standard (see discussion in Kahneman & Tversky 1982). Some of the attempts to vindicate the base rate fallacy have, unfortunately, fallen in this category. The ignoring of base rate considerations is a common phenomenon, and often unjustifiable. At the same time, intuition is often sensitive to considerations that are not captured by simplistic formal models. The astute judge under uncertainty would do well to rely on neither of them to the total exclusion of the other.

References

- Ajzen, I.: 1977. Intuitive theories of events and the effects of base rate information on prediction. In: *Journal of Personality and Social Psychology* 35, 303-314
- Bar-Hillel, M.: 1980. The base-rate fallacy in probability judgments. In: *Acta Psychologica* 44, 211-233
- Bar-Hillel, M.: 1981. Representativeness Reconsidered. *Decision Research Report* 81-1
- Bar-Hillel, M. & Fischhoff, B.: 1981. When do base rates affect predictions? In: *Journal of Personality and Social Psychology* 41, 671-680
- Beyth-Marom, R. & Arkes, H.R.: 1983. Being accurate but not necessarily Bayesian: Comment on Christensen-Szalanski and Beach. In: *Organizational Behavior and Human Performance*, in press
- Borgida, E. & Brekke, N.: 1980. The base rate fallacy in attribution and prediction. In: J.H. Harvey/W.J. Ickes & R.F. Kidd (Eds.), *New Directions in Attribution Research*, Vol. 3. Hillsdale, N.J.: Erlbaum
- Carroll, J.S. & Siegler, R.S.: 1977. Strategies for the use of base rate information. In: *Organizational Behavior and Human Performance* 19, 392-402
- Cohen, L.J.: 1979. On the psychology of prediction: Whose is the fallacy? In: *Cognition* 7, 385-407
- Cohen, L.J.: 1981. Can human irrationality be experimentally demonstrated? In: *The Behavioral and Brain Sciences* 4, 317-331
- Christensen-Szalanski, J.J.J. & Beach, L.R.: 1982. Experience and the base rate effect. In: *Organizational Behavior and Human Performance* 29, 270-278
- Dershowitz, A.: 1971. Imprisonment by judicial hunch. In: *American Bar Association Journal* 57, 560-564

- Eddy, D.M.: 1982. Probabilistic reasoning in clinical medicine: Problems and opportunities. In: D. Kahneman/P. Slovic & A. Tversky, *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press
- Estes, W.K.: 1976. The cognitive side of probability learning. In: *Psychological Review* 83, 37-64
- Fischhoff, B. & Bar-Hillel, M.: 1981. Diagnosticity and the base rate effect. In: *Decision Research Report* 81-9
- Fischhoff, B. & Bar-Hillel, M.: 1982. Focussing techniques as aids to inference. Unpublished manuscript
- Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. In: *Journal of Experimental Psychology: Human Perception and Performance* 8, 330-344
- Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1979. Subjective sensitivity analysis. In: *Organizational Behavior and Human Performance* 23, 339-359
- Foltz, A.M. & Kelsey, J.L.: 1978. The annual Pap test: A dubious policy success. In: *Milbank Memorial Quarterly/Health and Society* 56(4), 426-462
- Freedman, D./Pisani, R. & Purves, R.: 1978. *Statistics*. New York: W.W. Norton
- Ginosar, Z. & Trope, Y.: 1980. The effects of base rates and individuating information on judgments about another person. In: *Journal of Experimental Social Psychology* 16, 228-242
- Huff, D.: 1959. *How to take a chance*. Harmondsworth: Pelican Books
- Kahneman, D. & Tversky, A.: 1972a. Subjective probability: A judgment of representativeness. In: *Cognitive Psychology* 3, 430-454
- Kahneman, D. & Tversky, A.: 1972b. On prediction and judgment. In: *ORI Research Monograph* 12(4)
- Kahneman, D. & Tversky, A.: 1973. On the psychology of prediction. In: *Psychological Review* 80, 237-251
- Kahneman, D. & Tversky, A.: 1982. On the study of statistical intuitions. In: *Cognition* 11, 123-141
- Kassin, S.M.: 1979. Consensus information, prediction, and causal attribution: A review of the literature and issues. In: *Journal of Personality and Social Psychology* 37, 1966-1981
- Lykken, D.T.: 1975. The right way to use a lie detector. In: *Psychology Today* 8, 56-60
- Lyon, D. & Slovic, P.: 1976. Dominance of accuracy information and neglect of base rates in probability estimations. In: *Acta Psychologica* 40, 287-298
- Manis, M./Dovalina, I./Avis, N.E. & Cardoze, S.: 1980. Base rates can effect

- individual predictions. In: *Journal of Personality and Social Psychology* 38, 231-248
- Meehl, P.E. & Rosen, A.: 1955. Antecedent probability and the efficacy of psychometric signs, patterns or cutting scores. In: *Psychological Bulletin* 52, 194-216
- Nisbett, R.E. & Borgida, E.: 1975. Attribution and the psychology of prediction. In: *Journal of Personality and Social Psychology* 32, 932-943
- Nisbett, R.E./Borgida, E./Crandall, R. & Reed, H.: 1976. Popular induction: Information is not always informative. In: J.S. Carroll & J.W. Payne (Eds.), *Cognition and Social Behavior*
- Tversky, A. & Kahneman, D.: 1980. Causal schemas in judgment under uncertainty. In: M. Fishbein (Ed.), *Progress in Social Psychology*. Hillsdale, N.J.: Erlbaum
- Tversky, A. & Kahneman, D.: 1982a. Evidential impact of base rates. In: D. Kahneman/P. Slovic & A. Tversky, *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press
- Tversky, A. & Kahneman, D.: 1982b. Judgments of and by representativeness. In: D. Kahneman/P. Slovic & A. Tversky, *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press

This Page Intentionally Left Blank

THE TWO CAMPS ON RATIONALITY

HELMUT JUNGERMANN

*Technische Universität Berlin
Institut für Psychologie
F.R. Germany*

The status of the rationality concept

If one postulates that people are generally *rational*, one meets today usually objection or, at best, scepticism; biases, errors and faults are described and illustrated to prove that the postulate has little empirical justification. If one declares a person *irrational*, however, one equally will meet usually protest from this person; he or she will explain that there were good reasons for the judgment or decisions questioned. The contradiction might be due to the words "generally" and "usually" in the above statements: People are not *always* rational and people will not *always* claim to be rational. However, the present debate about the quality of human judgment and decision indicates that the controversy is not only about generalizations and exceptions (e.g., Einhorn and Hogarth, 1981; Nisbett and Ross, 1980; Cohen, 1979; Kahneman and Tversky, 1982; Edwards, 1983; Fischhoff, 1983; Berkeley and Humphreys, 1982; Phillips, 1983). The participants in this debate tend to avoid the term "rationality" because it is overloaded with many connotations, but just for that reason I find it useful as an umbrella under which most contributions fit fairly well; they all treat one or another definition or meaning of rationality.

Rationality is not a genuine term of scientific psychology but rather a concept of philosophy and economics. The most common, and in this context most relevant, definition says that an action is rational if it is in line with the values and beliefs of the individual concerned; or more precisely, if it is "logical" or "consistent" as stated in a set of axioms. This definition specifies rational behavior normatively. Empirical research can study whether actual human behavior is rational in the sense that it obeys the norm.

To measure human behavior with a normative yardstick is not the rule but rather the exception in cognitive psychology. In the psychology of *perception*, for instance, the physical world is taken as a standard when perceptual illusions are investigated. Normally, however, the physical stimuli are used to provoke responses; one tries to understand and explain these responses and they are not evaluated as 'deviant from' or 'consistent with' something. The psychology of *language* worked for some time on the assumption that human language could be evaluated with the linguistic yardstick of an ideal speaker or hearer. But this assumption turned out to be not very helpful in understanding actual human language, and research began to focus on psychological models like semantic networks. Finally, the psychology of *thinking* used, and to some extent still uses, formal logic as a yardstick for the study of deductive reasoning. But not only is this merely a small area within the psychology of thinking, the approach itself has increasingly been disputed as mistaken. Similarly to the development in psycholinguistics, the focus has been directed increasingly on the content rather than on the formal characteristics of thinking. In all these areas of research, the use of a physical or logical yardstick plays only a minor role today or has been abandoned altogether.

In the psychology of *judgment and decision*, on the other hand, normative models have been the most important research tools since its beginnings in the late fifties. Research has been primarily concerned with studying, explaining and interpreting discrepancies between predictions derived from normative models and actual judgments

and decisions. The most prominent models of judgment are probably Bayes' theorem and the multiattribute utility models, and the normative model of decision is the SEU model. Most research is still oriented towards these models and the debate about the quality of human judgment and decision is still centered around the idea of rationality embodied in these models.

In a somewhat exaggerated manner, I will distinguish two camps in this debate, one that points to the deficiency and one that argues for the efficiency of human judgment and decision. The *pessimists*, as I will call the members of the first camp, claim that judgment and decision making under uncertainty often show systematic and serious errors, due to in-built characteristics of the human cognitive system. Violations of rationality, particularly of the SEU model, are interpreted as true deficits of the decision maker. The *optimists* of the other camp claim that judgment and decision are highly efficient and functional even in complex situations. Observed violations of rationality axioms are interpreted as unjustified evaluations based on inappropriate theoretical assumptions or empirical approaches on the part of the researcher. In the following, I will describe the approach to rationality taken in both camps, each of which has various factions, and discuss some points that make the debate so complicated and, sometimes, obscure. The description cannot be as differentiated as the approaches are, of course, but I hope that it is essentially correct; the interested reader is referred to the original sources.

The pessimists: Biases are in people

Since Simon (1955) proposed the concept of "bounded rationality", the strongest attacks against human rationality came from Tversky and Kahneman (e.g., 1974), Slovic (1972), Janis and Mann (1977), and Nisbett and Ross (1980). The general tendency of this research was that human judgment and decision making ability and capacity is indeed limited, leading to violations of rationality principles. There are three variants in this camp in explaining these violations: As results

of *judgmental biases*, of *representational faults*, and of *coping defects*.

Judgmental biases: The most extensive and influential work of interest in this context has been the research of Tversky & Kahneman on probabilistic thinking. While in the sixties, Peterson & Beach (1967) in their review came to the conclusion that people form and revise their beliefs according to the normative principles of statistics. Tversky & Kahneman (e.g., 1974) offered a different conception: Judgments of probabilities are often severely biased because people in many situations rely on heuristics which, although generally efficient, can sometimes lead to systematic errors. For example, people often judge the probability of an event according to the *representativeness* of the event for the underlying population or for the generating process. However, "this approach to the judgment of probability leads to serious errors, because similarity or representativeness is not influenced by several factors that should affect judgments of probability" (Tversky & Kahneman, 1974, p.1124). Such factors are, for instance, base rates, sample size and the reliability of information. Another heuristic people use is the saliency or *availability* of information, i.e., how easily instances of the event whose probability is to be assessed can be retrieved from memory. Since ease of retrieval is also influenced by other factors than the actual frequency of the event (e.g., recent occurrence) the use of this heuristic can likewise result in systematic judgmental errors. Other heuristics described in this research are "anchoring and adjustment" and "simulation" (for a collection of the most important papers on biases and heuristics see the recent book by Kahneman, Slovic and Tversky, 1982).

Judgmental biases may produce inconsistent decisions. Such inconsistencies would then not reflect violations of assumptions of the model (e.g., SEU model) but rather incorrect input. However, the discussion has mostly been restricted to the quality of the judgment itself. The judgment is compared to a standard like relative frequency or the inference based on a statistical model for which the researcher has defined the relevant data (e.g., what the base rates are). In both cases, deficiency means deviation of the judgment from some

"objective" quantity, and the "objectivity" is rooted in the real world, or rather the researcher's view of the real world. The implicit notion of rationality, then, is not consistency but realism; it is less *formal* rationality that is questioned than *substantive* rationality: People use cognitive inference and retrieval strategies which often lead to (in the researcher's view) substantively incorrect judgments about the world.

Implicitly, this meaning of rationality pervades the discussion on biases and heuristics in probabilistic thinking. It is interesting to find it made *explicit* with respect to judgments which are equally important for decision making but much less debated, namely, utility judgments (March 1978). Tversky & Kahneman (1981) question the rationality of utility judgments that turn out to have been unrealistic anticipations of satisfaction at the time when the consequences actually incur. But biases and heuristics in the judgment of utility are still unexplored, maybe due to the lack of a yardstick like relative frequency it offers for probability judgments. A different approach is needed here, maybe taking the work of Ainslie (1976) on preferences as a function of time delay of satisfaction and the economic approach to 'rational expectations' as a starting point. Rationality in this sense, i.e., as realism, certainly comes close to the meaning of this term in common sense: We often call hopes and fears "irrational" which we consider as extremely unrealistic. The question, of course, is who is to define reality.

Representational faults: A more recent argument against rationality, raised by the same group of researchers, concerns the effect of different frames for the decision problem on people's decision making behavior (Tversky and Kahneman, 1981; cp. also Payne, 1982). A striking example is the experiment in which a majority of subjects behaved risk averse when they had to decide between two options that were formulated in terms of lives saved, but behaved risk seeking when the options were formulated in terms of lives lost - although the (expected) values of the options were in both cases identical. This violation of rationality, i.e., of the SEV model, is interpreted by the assumption that

people code the possible outcomes as gains and losses rather than as final states, and the framing of the options induces in one case coding in terms of gains and in the other case coding in terms of losses, i.e., it induces subjects to focus on different parts of their utility functions. Since the utility functions are different below and above the reference point, apparently inconsistent preferences result. Coding outcomes in terms of gains and losses is, according to Kahneman and Tversky (1979), only one of several cognitive mechanisms which people use to edit, or represent, decision problems before the options will be evaluated, and these operations can also lead to violations of rationality. Other editing mechanisms are, for instance, the segregation of riskless components of prospects and the cancellation of components which are shared by two prospects.

The representational errors resulting from the application of such mechanisms are likened to perceptual illusions (Tversky and Kahneman, 1981). This implies the assumption that there exists one and only one correct representation of the problem, as there exists only one veridical representation of the physical world. Inconsistent preferences are understood as results of a deficient perception and interpretation of the decision problem.

It is interesting to note that with this argument, the context for discussing rationality is being enlarged. Previously, the domain of theoretical and empirical research was judgment and evaluation only, but *not* the cognitive activities which forego these steps, e.g., how the problem structure is generated. Only recently, interest has shifted to these early phases of the decision making process (e.g., Pitz, Sachs and Heerboth, 1981; Jungermann, von Ulardt and Hausmann, 1983), but this research has not been generally linked to the discussion about the deficiency/efficiency of judgment and decision.

Coping defects: The third attack on rationality comes from Janis and Mann (1977) who look at people's decision making behavior from a motivational perspective. They distinguish various coping patterns that people use in handling stress of decision situations, and only one of

them corresponds to the rational behavior as explicated in decision theory. The other four patterns lead to defective decision making. For example, if the coping pattern is "defensive avoidance", the person escapes the decisional conflict by procrastinating, shifting responsibility to someone else, or constructing wishful rationalizations to bolster the least objectionable alternative, remaining selectively inattentive to corrective information. Although these patterns, as Mann and Janis (1982) concede, can occasionally be adaptive in saving time and effort, they often lead to defective decision making.

"Deficiency" is here not directly defined in relation to some normative model, e.g., the SEU model, but more generally by the failure of people "to assimilate and combine information relating to outcome expectations and values" (p.347) and particularly "to make use of the resources available to them for engaging in effective search for and appraisal of alternatives - within the limits of their cognitive capabilities and within limits imposed by powerful social constraints" (p.346). In a large number of experimental studies, Janis and Mann (1977) have collected evidence for these coping defects, and they have tried to identify the conditions under which people exhibit different types of information processing behavior in decision situations.

Thus, Janis and Mann also emphasize the deficiencies of human judgment and decision, but in their perspective, the sources are motivational, not cognitive in nature: "We see man not as a cold fish, but as a warm-blooded mammal, not as a rational calculator always ready to work out the best solution but as a reluctant decision maker - beset by conflicts, doubts, and worry, struggling with incongruous longings, antipathies, and loyalties ..." (1977, p.15). This latter statement illustrates that, although the focus is explicitly on decision making under stress, the authors tend to generalize their theory to human decision making in general.

A similar approach, though from a cognitive standpoint, has been taken by Dörner who examined how people operate in highly complex and dynamic situations (e.g., Dörner, 1983; Dörner, Kreuzig, Reither and Stäudel, 1983). In one of the studies, for instance, he used a computer

model of a little town and had the subjects reign this town as mayor for ten simulated years in interaction with the computer. Having to cope with such ill-defined problems, subjects often apparently use a number of heuristic procedures which might be considered deficient. Dörner et al. do not compare the observed behavior with normative models of judgment and decision, however, but rather check whether the subjects' behavior meets certain demands of the situation. For example, requirements resulting from characteristics of the aspired goal are often managed by "redefining" or "forgetting" the final goal, by thematic vagabonding (e.g., swinging from one area of pursuit to another), or by encapsulation (e.g., sticking obstinately to a theme irrespective of the changing situation). Although the deficiencies displayed in these experiments are probably largely due to cognitive overload, some of the heuristics might well be used also in less complex situations.

In summary, the members of this camp view human judgment and decision making as deficient in several respects: Judgments are sometimes systematically biased, due to the use of heuristics; decisions are sometimes inconsistent, due to errors in the representation of the problem; and information search and combination is often defective, due to motivational factors. Mostly, this deficiency is not seen simply as a consequence of cognitive overload in highly complex or unfamiliar situations, but as rooted in mechanisms working within the human information processing system itself. People are prone to violate principles of rationality. A key assumption of at least the cognitive variants described is that there is a reliable and valid yardstick for the evaluation of judgment and decision, namely, the objective reality, and consequently that there is some kind of objectively veridical mode of information processing which is mapped in the respective normative model.

The optimists: Biases are in research

In the last few years, a kind of counter-movement has developed. The members of this new camp partly question the validity of the other camp's findings, and partly emphasize the implicit rationality of human judgment and decision behavior. The biases, they argue, are not in human behavior but in the analysis of this behavior in the other camp (Berkeley and Humphreys, 1982). Interestingly enough, a leading representative of this camp has been a founder of the other camp with his research on conservatism in human information processing (Edwards, 1968); other researchers I include in this camp are Beach and Mitchell (1978), Einhorn and Hogarth (1981), Berkeley and Humphreys (1982), and Phillips (1983). Three theoretical arguments are raised in particular, which I will call the *meta-rationality* argument, the *continuity* argument and the *structure* argument.

The meta-rationality argument: The essence of this argument is that decision behavior which violates principles of rationality as, for instance, the principle of maximizing subjectively expected utility, can be described as perfectly rational if the cognitive *costs* of being rational are taken into account. To illustrate the point, when somebody wants to buy a book at the train station for a long ride, he will probably not check all available books in order to find the best of all but he will look at a few and then buy the first that he finds reasonably attractive. This person can be described as working on the satisficing principle, i.e., as not behaving fully rational from a SEU model perspective. But to many people, that is counter-intuitive. They would probably call a person non-rational if he *would* check all books in the store. With finite time and resources available, it is not rational to spend infinite effort on the exploration of all potential consequences of all options. Rather, the decision costs are weighted against the potential benefits resulting from the application of a decision strategy, and this may lead to violations of SEU model rationality which are, however, perfectly rational.

This point has been made by various authors (e.g., Miller and Starr, 1967; Einhorn and Hogarth, 1981; Janis and Mann, 1977; Hogarth, 1980; Montgomery and Svenson, 1976; cp. Payne, 1982). It is most explicitly and clearly represented in the contingency model of Beach and Mitchell (1978). In this model it is assumed that people make meta-decisions between the various decision strategies they have in their repertoire. Strategies differ with respect to the cognitive effort required and the probability with which they lead to an optimal solution (e.g., SEU maximizing, satisficing, elimination-by-aspects, coin flipping). Strategy selection is seen as contingent upon a (cost/benefit) compromise between the decision maker's desire to make the best decision and his or her negative feelings about investing time and effort in the decision making process. The strategy that is perceived as yielding the maximum net gain is the one selected. The specific choice depends on characteristics of the situation (e.g., familiarity of the problem, time pressure) and of the subject (e.g., knowledge, ability). Various hypotheses derived from this model have been experimentally investigated and have generally been confirmed (e.g., Christensen-Szalanski, 1978; McAllister, Mitchell and Beach, 1979; Christensen-Szalanski, 1980).

The rationality cost, i.e., the cognitive effort associated with each strategy, is not formalized in Beach and Mitchell's (1978) model. For the comparison of multidimensional alternatives, Johnson (1979) has proposed a process model which allows the computation of the number of mental operations required. A similar method has been proposed by Shugan (1980). In both models it is assumed, however, that the researcher knows the number of attributes. For more simple strategies, a formal approach might be more difficult.

In the model of Beach and Mitchell (1978), maximizing of subjectively expected utility is only one of several available strategies each of which may be chosen rationally as the best strategy under the given conditions, even coin flipping. Violations of 'classical' rationality are considered as errors that the subject anticipates and, more important, tolerates. People are global maximizers with local inconsis-

tencies (Elster, 1979). Renunciation of cognitive effort might not only result, however, from the calculation that the costs outweigh the expected benefits but also from the knowledge that "*the very act of deliberating* can modify the character for the worse, and in ways judged even more important, through the stultifying effects on spontaneity" (Elster, 1979, p.40). Thus, to save spontaneity, which in itself has a value, a person might rationally select a strategy possibly leading to violations of SEU rationality.

The continuity argument: The core of this argument is the conceptualization of judgment and decision as moments in a continuous process. They sometimes may appear biased or deficient if treated or tested as discrete events, while they might in fact be very functional when considered as moments in a continuous and changing environment (Hogarth, 1981). An experiment by Ronen (1973), described in Hogarth (1981), can serve as an illustration: Subjects had to choose between two options with identical (positive) payoffs and probabilities of success. The outcomes were dependent on the results of two sequential events. In one option, there was a higher probability of success in the first step and a lower probability in the second step, while it was the other way around with the alternative option. Since the SEUs of the two options were equal, rationality would imply indifference. However, subjects preferred the option with the higher probability in the first step. Hogarth (1981) provides the following explanation: If people are used to a changing environment, they might give less weight to the probability of the second step since the situation might have changed when this phase has been reached; it might be attractive to stay in the game as long as possible, and therefore to choose the option with the higher probability of success in the first phase.

More generally, many experiments on probabilistic information processing have excluded the possibility of feedback and redundancy which, however, is characteristic of the environment and thus relevant for judgment and decision. Models have been tested under the assumption of stable environment, while subjects might have operated under

the assumption of a changing environment. A number of biases, demonstrated with the discrete approach, can be interpreted as indicators of cognitive mechanisms and strategies which are actually very functional in a continuous environment. Hogarth (1981) discusses several assumptions underlying prescriptive models which explicitly do not take this continuity into account, e.g., the *existence of a stable time horizon*: An example given by Tversky and Kahneman (1973) for the use of availability resulting in a biased judgment is re-interpreted by Hogarth (p.206) with the argument that in the specific experimental situation under time pressure, the application of this heuristic could be considered as very functional: People use the cues which are most available first, because in normal life one can expect to have a chance to correct such a judgment. It is therefore important to specify under which conditions the application of a heuristic is functional and under which it is not. Another assumption discussed is the *stability of preferences and goals* which maybe wrong for various reasons (March, 1978): First, preferences develop over time. Choices imply, however, anticipations of future preferences, i.e., preferences at the time when the consequences incur (a point I discussed earlier in Section 2). Secondly, preferences are formed by experiences, and these experiences are often sought actively by people, sometimes in order to know more about their preferences. Thirdly, preferences are usually characterized by a high degree of ambiguity. Such ambiguity can be functional, because the person is open for further information and saves mental energy to clarify the ambiguity. Other assumptions discussed are the *stationarity of probabilistic processes*, the *independence of judgmental effects on consequences*, and the *abstraction from the competition character of action*. The continuity argument is raised by Hogarth (1981) not only against the contention of judgmental biases but also against the contention of representational errors.

The attack on the pessimists' camp can take two forms. Demonstrations of deficient judgment and decision can either be attributed to inadequate assumptions of the model, like fixed time horizons; the observed behavior was in fact efficient and functional, the model

was wrong. Or, one can claim that the experimental situations used for testing the model had a very low ecological validity and that therefore biases and errors were possibly artifacts, encouraged by the researcher.

The structure argument: The third argument against the other camp has been raised in particular by Berkeley and Humphreys (1982) and Phillips (1983). The claim is that demonstrations of cognitive deficiency are questionable due to the neglect of the subjects' internal structural representation of the problem. The characterization of the observed behavior as biased rests on the assumption that the subjects share the experimenter's understanding of the problem structure. A common understanding of the problem structure can be assumed to be established by the use of process tracing methods or through the experimental instructions; or, it can be assumed to exist a priori because of the objective nature of the task, as illustrated by Tversky and Kahneman (1981) in their use of the analogy of perception.

As an example, Berkeley and Humphreys (1982) discuss the problem presented to subjects by Tversky and Kahneman (1981) in which two options are formulated either in terms of lives lost or in terms of lives saved. In the first case, subjects behaved risk averse, in the second case they behaved risk prone, interpreted by Tversky and Kahneman as inconsistent preferences since "it is easy to see that the two problems are effectively identical" (p.453). Berkeley and Humphreys question this interpretation because the "uncertainty, concerning human agency in affecting subsequent states of the world, is left unresolved" (p.222) in Tversky and Kahneman's formulation of the problem. Berkeley and Humphreys offer a different structural representation the subjects might have developed in order to resolve this uncertainty which would result in preferences perfectly consistent with SEU theory. A similar reasoning is found in Hogarth (1981) and Phillips (1983).

The second approach, in which the problem structure is imposed under the assumption of a common understanding, does not investigate experimental data in their own right as products of the subjects' reasoning, but compares them with data resulting from the application

of a normative model to the problem in question, "the assumption being that the model represents the basis for rational choice (or inference) in the presented problem" (Berkeley and Humphreys, 1982, p.230). Two implications of this assumptions are questioned by the authors: First, the naturalization of small worlds, which means the exclusion of the subjects' individual "goal-closing" of his or her large world into the small world of the experimental moment and presenting the observed departures from veridicality as *cognitive* rather than motivational biases (p.233). Second, the utilization of normative models as ideal types, i.e., as standards of comparison with the intuitive model of the subject; this requires, however, the investigation of whether there is a common understanding of the way in which the problem universe is to be goal-closed and thus does not permit imposition of a problem structure (p.234).

This argument, then, focuses on the structural representation of the task and its significance for the interpretation of judgments and decisions. The representation depends strongly, however, on the context within which it is developed. Different contexts evoke different knowledge - elements and structures of knowledge - and thus will often lead to different behaviors. Each behavior might be consistent within the given context although inconsistent with some behavior towards the same task in a different context. The description of behavior as deficient or non-rational is not justified if it is not invariant over multiple contextes; to the contrary, invariant behavior would be deficient.

Two views are still possible here: One can go as far as Phillips (1983) and argue phenomenologically that there is no criterion for defining some "objective" problem representation as implied in Tversky and Kahneman's (1981) analogy between judgmental errors and perceptual illusions. Or one can assume such a criterion but conclude that "task representation may be of more importance in defining errors than the rules they (the people) use within that representation" (Einhorn and Hogarth, 1981, p.60), e.g., if one considers the case of a paranoid person. In both views, however, the problem of structure is essential

for a discussion of deficiency/efficiency.

In summary, the members of this camp challenge the view that human judgment and decision is cognitively deficient. They question this conclusion with different arguments: Because an important parameter has been neglected, namely, the cost of a decision strategy; because judgments and decisions are treated as discrete events and not as moments in a continuous process; and because no attention has been paid to the internal structural representation of the problem as the person- and context-dependent basis of judgment and decision. The emphasis in all three positions is less on demonstrating efficiency of human behavior than on studying more carefully under what conditions people show which kind of behavior, and not "too easily to adopt a crude view of human rationality" (Slovic, 1972).

Resumée and Outlook

It seems that the history of research on human judgment and decision making in the last thirty years represents a typical example of scientific progress. Against the once dominant model of rational man, the camp I have called pessimistic formed and questioned this conception by demonstrating deficiencies in judgment and decision. Now after this camp has ruled decision research for many years, it comes under attack by a new group that I have called optimistic which points to weaknesses in the former concepts and suggests even more differentiated perspectives of judgment and decision. Both camps have gained their impetus from the critique of the respective prevailing conception. Some of the criticisms, particularly of the optimists's camp against the pessimists' camp, have already been mentioned in the previous section. However, I will briefly sum up these points and add those more general arguments which have been exchanged between the two camps more recently.

The critique of the pessimists' position focuses on the following points: First, the experiments are of low ecological validity in the sense that they do not represent the majority of situations in which

people have to give judgments and make decisions. Second, the experiments were designed such that biases and errors are not surprising since the situations and problems were taken out of any real life context. Third, important parameters like cognitive effort have not been accounted for in the models used as yardsticks. Fourth, whether subjects cognitively represented the problem in the way the researcher assumed is rarely investigated. A more general, fifth argument is that this research has focused too much on errors rather than on the cognitive processes per se; not only is it hard to define what errors are, they also constitute atypical behavior samples.

The rejoinder of the pessimists' camp is: First, meta-theories on the selection of strategies according to cognitive effort considerations are methodologically more or less immune to charges of "irrationality"; any behavior that violates the SEU model can be "rationalized" by recurring to some higher level of rationality. Second, demonstrations of differences between the experimental and real world situations do not prove that these differences matter. Third, speculations that subjects might have represented a problem cognitively differently than the researchers do not prove that they have done so. Fourth, the focus on errors is justified because (a) "they expose some of our intellectual limitations and suggest ways of improving the quality of our thinking", (b) they "often reveal the psychological processes and the heuristic procedures that govern judgment and inference", and (c) they "help the mapping of human intuitions by indicating which principles of statistics or logic are non-intuitive or counter-intuitive" (Kahneman and Tversky, 1982, p.124). Clearly, these are defensive arguments but therefore not less reasonable than the offensive arguments of the authors. It is not to be expected that this camp will capitulate as easy under the attack as did the "rational man" when that concept was besieged by today's defenders.

As Heraclit said, war is the father of all and the king of all. The theoretical and empirical work in both camps as well as the dispute between them has generated many data, insights and ideas. The understanding of judgment and decision making has been dependend and

differentiated and has opened new, more comprehensive research perspectives. I will characterize briefly some essential aspects of the present state of the debate and of the direction future research might take.

For almost all arguments, the theoretical point of reference is still the SEU model or some variant of it (e.g., prospect theory), despite some harsh critique (e.g., Fischhoff, Goitein and Shapira, 1982). While one side provides evidence that people do not behave as the model predicts, the other side demonstrates that with different theoretical assumptions or experimental settings the same behavior might not violate the model at all. Pitz said (in 1977) that "models based upon some variant of expected value theory have the attraction of being familiar and easy to work with, and inertia is one of the strongest forces in nature" (p.421); another reason is probably that the SEU model is hard to falsify since "with sufficient ingenuity, one can always find something that a particular decision maker has maximized in a particular situation" (Fischhoff et al., 1982, p.317). Pitz also said that "until a well developed, systematic alternative to the normative model is available, it is likely that new approaches will have no more lasting success than did Gestalt psychology" (1977, p.421). This still seems to be a fairly valid description of the present situation. However, that no real alternative has been offered might also be taken as an indicator of the strenght of the model as a core of theories of judgment and decision. The wide use of expectancy-value models also in other areas of psychology, particularly in motivation research, is a further indicator (cp. Feather, 1982).

The issue is not anymore what the limitations of the human cognitive system are. All approaches assume that there are boundaries for rationality in situations of cognitive overload due to the capacity and processing limits of the system. The issue is how people form judgments and make decisions, and particularly whether they operate rationally *within* the constraints. While one side focused on errors, biases and fallacies, the other side considers this a theoretically questionable and also too negative approach. In this view, research on judgment and decision making has been driven too much by a concern for errors rela-

tive to a normative standard the validity of which one can doubt with good arguments (Einhorn and Hogarth, 1981).

The focus on errors has recently been defended by Kahneman and Tversky (1982) by pointing towards various advantages of this approach (see above). Kahneman and Tversky also argue that "the emphasis on the study of errors is characteristic of research on human judgment, but it is not unique to this domain: we use illusions to understand the principles of normal perception and we learn about memory by studying forgetting" (1982a, p.123). I don't quite agree, for two reasons: First, because, although illusions are a subject of interest in the psychology of perception, the *emphasis* has certainly not been on errors. Second, memory research is not interested in errors *as such* but uses them as a dependent variable for testing models about storing or retrieval of information, i.e., about the "normal" memory processes and structures; researchers on memory do not demonstrate errors in their research, as do many researchers on judgment and decision. However, the recent discussion of the issue seems to have led to some rapprochement of positions. Kahneman and Tversky concede that "although errors of judgment are but a method by which some cognitive processes are studied, the method has become a significant part of the message" (1982a, p.124), and they accentuate stronger than before, e.g., that the notion of judgmental heuristics should "provide a common account for both correct and incorrect judgments" (1982b, p.325). If it is agreed, then, that "theoretically it is necessary to understand *all* reasoning data, regardless of their conformity to a normative rule system" (Evans, 1982, p.319), the study of errors becomes a question of research strategy. Kahneman and Tversky, in their reply to Evans, argue that "errors may sometimes be more informative than correct judgments, which can be produced *either* by an initial valid intuition *or* by a subsequent correction" (1982b, p.325).

Another important progress is to be seen in the now generally accepted conceptualization of judgment and decision as parts of a multi-stage cognitive process. Actually, activities that precede the "moment of decision" seem presently to be the prevailing object of theoretical

and empirical work - in particular the acquisition and processing of information, or, in other terms, the generating and structuring of knowledge. Some work on predecisional cognitive processes is already available (cp. Einhorn and Hogarth, 1981), but new questions have come up. For instance, what are the implications of using the "conversational paradigm" (Berkeley and Humphreys, 1982; Kahneman and Tversky, 1982a)? How can we elicit or infer the cognitive representation of problems? We need theories of problem representation that explain the use and the results of particular judgmental and decision making strategies. - There is also increasing interest in post-decisional processes, particularly the implementation of decisions (cp. Gasparski, 1980). Again, research has focused primarily on "errors", i.e., on the phenomenon that people often do not implement their mental decisions due to motivational and emotional factors (e.g., Sjöberg, 1980; Elster, 1979). Elster has analyzed the various strategies that people use to overcome this "weakness of will", interfering between decision and action (using Ulysses as a prominent example who bound himself to the mast in order not to give in to the temptation by the sirens. He put this problem in the context of the rationality debate: "Man often is not rational, and rather exhibits *weakness of will*. Even when not rational, man knows that he is irrational and can *bind himself* against the irrationality. This second-best or imperfect rationality takes care both of reason and passion. What is lost, perhaps, is the sense of adventure" (Elster, 1979, p.111).

A final lesson to be learned from the debate might be that one should avoid the term rationality in psychology at all. Obviously, the concept is used with different meanings; it is certainly not longer defined exclusively by formal coherence and consistency. An alternative way of handling this difficulty is, however, to distinguish explicitly various meanings of the term. Besides formal rationality, concepts like substantive rationality and procedural rationality might be useful for a theory of judgment and decision (Simon, 1978). *Substantive* rationality captures an important aspect of the term as it is used in common language, namely, how realistic, correct, adequate some judgment or decision is with respect to the real world - the paranoid with his

perfectly consistent beliefs and values provides an example. Rationality is not only a question of whether a choice is in line with a person's belief and preferences, but also a question of what sort of preferences and beliefs the person holds, a point elaborated also by Einhorn and Hogarth (1981). Differently stated, a decision might be consistent (and thus formally rational), but the judgments that provide the input for the decision might be very poor (and thus the choice might be substantively not rational). *Procedural* rationality can be linked to substantive rationality easily if one asks what information is searched for and used by a person to form values and beliefs, whether the person tried hard enough to anticipate future consequences of current actions and future preferences for those consequences, etc. What Janis and Mann (1977) describe as "vigilant information processing" might be considered one aspect of this kind of rationality. Particularly in a social context, when decisions have to be justified, procedural rationality can be much more important than substantive matters about which agreement often cannot be reached.

My conclusion from the debate between the two camps with respect to the rationality problem is threefold: *First*, we should be liberal (which does not mean vague) in our use of the rationality concept. The different meanings that I have briefly described (and others, cp. March, 1978) are all useful tools in analyzing judgment and decision processes. *Second*, we may use the concept in its prescriptive sense legitimately in situations where prescription is asked for, like in decision aiding or in decision training. Criteria for defining rationality in a prescriptive sense can be taken from our knowledge about the real world, from our intellectual system and from the social consensus - however relative these criteria are in a historical or cultural sense. *Third*, we should use the concept in descriptive research more cautiously because it too easily entices to label all behavior that does not meet the criteria as deviant or deficient, as errors, biases, or whatever. I subscribe to Kahneman and Tversky's balanced advice that we "should avoid overly strict interpretations, which treat reasonable answers as errors, as well as to overly charitable interpretations, which attempt to rationalize every

response" (1982a, p.124). *Finally*, the recognition of the conditionality of normative models on assumptions about the environment as well as the cognitive representation of that environment, most convincingly elaborated by Einhorn and Hogarth (1981) and Payne (1982), is in my opinion the synthesis emerging from the debate between the (pessimistic) thesis and the (optimistic) antithesis. Implicitly or explicitly, it is agreed to by members of both camps.

References

- Ainslie, G.: 1975. Specious reward. In: Psychological Bulletin 82, 463-496
- Beach, L.R. & Mitchell, T.R.: 1978. A contingency model for the selection of decision strategies. In: Academy of Management Review 3, 439-449
- Berkeley, D. & Humphreys, P.: 1982. Structuring decision problems and the 'bias heuristic'. In: Acta Psychologica 50, 201-252
- Christensen-Szalanski, J.J.J.: 1978. Problem-solving strategies: A selection mechanism, some implications, and some data. In: Organizational Behavior and Human Performance 22, 307-323
- Christensen-Szalanski, J.J.J.: 1980. A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. In: Organizational Behavior and Human Performance 25, 107-122
- Cohen, L.J.: 1979. On the psychology of prediction: Whose is the fallacy? In: Cognition 7, 385-407
- Dörner, D.: 1983. Heuristic and cognition in complex systems. In: R.Groner, M.Groner & W.F.Bischof (Eds.), Methods of heuristics. Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- Dörner, D., Kreuzig, H.W., Reither, F. & Stäudel, T. (Hrsg.): 1983. Lohhausen. Bern/Stuttgart/Wien: Hans Huber
- Edwards, W.: 1968. Conservatism in human information processing. In: B.Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley & Sons
- Edwards, W.: 1983. Human cognitive capabilities, representativeness, and ground rules for research. In: P.C.Humphreys, O.Svenson & A.Vari (Eds.), Analysing and aiding decision processes. Amsterdam: North-Holland
- Einhorn, H.J. & Hogarth, R.M.: 1981. Behavioral decision theory: Processes of judgment and choice. In: Annual Review of Psychology 32, 53-88
- Elster, J.: 1979. Ulysses and the sirens. Cambridge: Cambridge University

Press, and Paris: Editions de la Maison des Sciences de l'Homme

- Evans, J.St.B.T.: 1982. On statistical intuitions and inferential rules: A discussion of Kahneman and Tversky. In: *Cognition* 12, 319-323
- Feather, N.T. (Ed.): 1982. *Expectations and actions*. Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- Fischhoff, B.: 1983. Reconstructive criticism. In: P.C.Humphreys, O.Svenson & A.Vari (Eds.), *Analysing and aiding decision processes*. Amsterdam: North-Holland
- Fischhoff, B., Goitein, B. & Shapira, Z.: 1982. The experienced utility of expected utility approaches. In: N.T.Feather (Ed.), *Expectations and actions*. Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- Gasparski, W. (Ed.): 1980. *Decision making and action*. Report. Warsaw: Polish Academy of Sciences
- Hogarth, R.M.: 1981. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. In: *Psychological Bulletin* 90, 197-217
- Janis, I.L. & Mann, L.: 1977. *Decision making*. New York: The Free Press
- Johnson, E.J.: In press. Deciding how to decide: The effort of making a decision. In: L.Sjöberg, J.A.Wise & T.Tyszka (Eds.), *Human decision making*. Lund: Doxa
- Jungermann, H., von Ulardt, I. & Hausmann, L.: 1983. The role of the goal for generating actions. In: P.C.Humphreys, O.Svenson & A.Vari (Eds.), *Analysing and aiding decision processes*. Amsterdam: North-Holland
- Kahneman, D., Slovic, P. & Tversky, A. (Eds.): 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press
- Kahneman, D. & Tversky, A.: 1979. Prospect theory: An analysis of decision under risk. In: *Econometrica* 47, 263-291
- Kahneman, D. & Tversky, A.: 1982a. On the study of statistical intuitions. In: *Cognition* 11, 123-141
- Kahneman, D. & Tversky, A.: 1982b. A reply to Evans. In: *Cognition* 12, 325-326.
- Mann, L. & Janis, I.L.: 1982. Conflict theory of decision making and the expectancy-value approach. In: N.T.Feather (Ed.), *Expectations and actions*. Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- March, J.G.: 1978. Bounded rationality, ambiguity, and the engineering of choice. In: *The Bell Journal of Economics* 9, 587-608
- McAllister, D.W., Mitchell, T.R. & Beach, L.R.: 1979. The contingency model

- for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. In: *Organizational Behavior and Human Performance* 24, 228-244.
- Miller, D.W. & Starr, M.K.: 1967. *The structure of human decisions*. Englewood Cliffs, N.J.: Prentice-Hall
- Montgomery, H. & Svenson, O.: 1976. On decision rules and information processing strategies for choices among multiattribute alternatives. In: *Scandinavian Journal of Psychology* 17, 283-291
- Nisbett, R. & Ross, L.: 1980. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall
- Payne, J.W.: 1982. Contingent decision behavior. In: *Psychological Bulletin* 92, 382-402
- Peterson, C.R. & Beach, L.R.: 1967. Man as an intuitive statistician. In: *Psychological Bulletin* 68, 29-46.
- Phillips, L.D.: 1983. A theoretical perspective on heuristics and biases in probabilistic thinking. In: P.C.Humphreys, O.Svenson & A.Vari (Eds.), *Analysing and aiding decision processes*. Amsterdam: North-Holland
- Pitz, G.: 1977. Decision making and cognition. In: H.Jungermann & G.de Zeeuw (Eds.), *Decision making and change in human affairs*. Dordrecht/Boston: D.Reidel
- Pitz, G.F., Sachs, N.J. & Heerboth, J.: 1980. Procedures for eliciting choices in the analysis of individual decisions. In: *Organizational Behavior and Human Performance* 26, 396-408
- Ronen, J.: 1973. Effects of some probability displays on choices. In: *Organizational Behavior and Human Performance* 9, 1-15
- Shugan, S.M.: 1980. The cost of thinking. In: *Journal of Consumer Research* 7, 99-111
- Simon, H.A.: 1955. A behavioral model of rational choice. In: *Quarterly Journal of Economics* 69, 99-118
- Simon, H.A.: 1978. Rationality as process and as product of thought. In: *American Economical Review* 68, 1-16
- Sjöberg, L.: 1980. Volitional problems in carrying through a difficult decision. In: *Acta Psychologica* 45, 123-132
- Slovic, P.: 1972. From Shakespeare to Simon: Speculations - and some evidence - about man's ability to process information. *Research Bulletin* 12, no.2. Eugene, Ore.: Oregon Research Institute
- Tversky, A. & Kahneman, D.: 1973. Availability: A heuristic for judging frequency and probability. In: *Cognitive Psychology* 5, 202-232

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

Tversky, A. & Kahneman, D.: 1981. The framing of decisions and the psychology of choice. In: *Science* 211, 453-458.

THE ROLE OF HEURISTICS IN MODELS OF DECISION

RUDOLF GRONER
MARINA GRONER & WALTER F. BISCHOF

*Universität Bern
Psychologisches Institut
Switzerland*

There is an increasing tendency for the word "heuristic" to become a fashionable label for a whole lot of quite different enterprises. Its meaning reaches from "imperfect solution method" up to "creative achievement". The purpose of this article is first to recapitulate the roots of this concept in philosophy and mathematics and its further development in psychology and Artificial Intelligence. Finally, some applications of the term in decision research are discussed.

1. Early History of Heuristics

The ethymological root of the word goes back to ancient Greece where it is the adjective associated with the verb *heuriskein* (= to find). The first heuristic methods have been used by Euclid and later described in detail by Pappus of Alexandria (for a more detailed historical review see Groner, Groner & Bischof, 1983). They distinguished between *Analysis* (assuming something to be true and working *backwards* to something already admitted to be true) and *Synthesis* (working *forwards* from already known material towards a new result). The two methods were mainly applied to constructions and proofs in the field of geometry.

In subsequent epochs, the history of heuristics became increasingly concerned with the development of algorithms, i.e. step-by-step procedures which automatically produce solutions to any problem out of a class of problems.

The term "algorithm" can be traced back to the name of a mathematician,

Mohammed ben Musa al Khovarezmi, who lived around the year 830 at Baghdad. He had written a book on rules of arithmetic developed by the Indians. In the Middle Ages it was translated into Latin under the name "Algoritmi de numero Indorum". Later, "algoritmi" was considered a Latin possessive case and "liber algoritmi" was used as a title for books on arithmetic. Finally when the origin of the word was forgotten completely, it was mistaken for a Greek word and the spelling was changed into "algorithmi". (Hankel, 1874).

The successful algorithmic methods in arithmetic and algebra stimulated Raymond Lullus, a Spanish philosopher of the thirteenth century, to search for a general algorithm that would produce *every* truth. In his famous work "Ars magna" he described a system where you start with a few classes of basic notions and by combining them in a systematic way you should invent everything. Lullus even invented a mechanical device to generate all combinations. He wrote the six classes, each containing nine concepts, on six concentric discs. Now by turning the discs against each other he could have got every possible combination. Lullus did not know combinatorics and probably underestimated the number of combinations his exhaustive search algorithm would generate, because he also thought of exploring all combinations of all pairs of basic concepts which would amount to $2^{176} \cdot 782 \cdot 336$ cases.

At later periods two eminent philosopher-mathematicians became involved in the area of heuristics from an algorithmic perspective: Descartes and Leibniz. Descartes believed that algorithms would eventually be found for all mathematical problems, perhaps also for the philosophical ones. He developed analytic geometry to transform geometrical problems into algebraic ones which can then be treated by algorithmic procedures. Changing the representation of a problem can be a way of solving it. The role which representational shifts play in the process of problem solving is an important part of the field of heuristics (Amarel, 1983).

Descartes (1908) recommended in his 21 heuristic rules to direct the mind ("regulae ad directionem ingenii") to reduce every problem, if possible, to algebraic equations. In particular, the rules say that one should first reduce a problem to its simplest form omitting all superfluous information.

one should try to quantify the problem in geometrical terms, drawing figures. One should determine the known and unknown magnitudes. Then, one should dissect the conditions that the unknowns have to satisfy in as many parts as there are unknowns in order to get as many equations, and finally one should reduce all equations to one. Solving a problem by means of algebraic equations is an example of the application of the method of analysis.

Descartes's remaining rules are more of a psychological nature and are concerned with the optimal functioning of the mind. They recommend to study a problem as long as it is necessary to understand it clearly, to use the senses, the imagination and the memory, to support the memory using external representations of the problem, and to practise the mind solving a lot of problems that have already been solved by others.

Leibniz criticized Descartes, saying that he had given us beautiful rules but no means to execute them. The rules were too general to be of any use. He even put his criticism into a sharp satirical form, summarizing Descartes's rules in the following statement: "Sume quod debes et operare ut debes, et habebis quod optas." (1880, vol.IV, p.329). ("Take what you have to take, and work the way you have to, and you will get what you are looking for".)

Leibniz himself took the greatest effort in searching for a general method which would solve every problem in an algorithmic way. In his youth, when he was studying geometry and algebra, the idea came to his mind that it should be possible to find a relatively small set of primitive concepts from which all known concepts and new ones could be derived (Couturat, 1961 b). Later, in his dissertation "De arte combinatoria", he worked out his plan. All notions should be decomposed into a few, simple, independent and non-contradictory elements. He called them "alphabet of human thoughts". These elements should be represented by appropriate symbols. Combining them in a systematic way would then lead to all known notions, and also generate new ones.

This gigantic plan involved several aspects on which Leibniz worked during his whole life. There was the analysis or reduction to primitives of the immense knowledge that had been accumulated, even then, by all sciences from mathematics to theology. There was the problem of choosing the right kind of

symbols to represent the primitives and their interrelations. Finally, a method had to be developed that would "mechanically" solve the fundamental problem of the art of discovery as Leibniz has seen it: that is, to find all true propositions in which a certain concept is either subject or predicate. That means this method should not only invent propositions but also decide about their truth. Of course, these tasks were not independent of each other.

In his search for primitive notions, Leibniz was influenced by the work of Lullus and by several of his own contemporaries who attempted to construct an universal language by defining some general categories as basic, representing them by symbols and combining these. One of them was Athanasius Kircher (1669) who had written a new version of the "ars magna". Although Leibniz esteemed those systems, he criticized that the number of basic categories had been arbitrarily chosen and some categories were not primitive and universal. He also stressed the importance of choosing a kind of representation of concepts that would facilitate thinking about them. In his opinion, the "characteristica universalis" was a decisive part of the art of discovery. The symbols representing the primitive notions should be, on the one hand as natural as possible, and on the other hand short and simple, such that they could be easily manipulated in calculation. The complex notions should be represented by some kind of combination of the symbols representing the primitive concepts. In this way, it should be possible to derive all properties of a complex notion from its representation. The only sample Leibniz has given of his "characteristica" is the representation of concepts of elementary geometry. (1880, vol. IV).

Leibniz has made several attempts at developing an universal system of representation, and with it, a method of derivation. Let us have a closer look at one dating from the sixteen-seventies. In this system Leibniz intended to use integers as symbols and the basic principle of arithmetic that every integer can be *uniquely* represented by a product of prime numbers. The primitive concepts are represented by prime numbers. The complex concepts which are conceived as combinations of primitive concepts, as we remember, are represented by products of the corresponding prime numbers. (1880, vol. VII).

Let us consider a hypothetical example. Assume that "animal" and "being rational" are primitive concepts. Then, they are represented by prime numbers, say 2 and 3, respectively. The concept "man" defined as rational animal is represented by 2×3 , which is 6.

Before we go on, we like to mention, by the way, that Leibniz who thought of using his system also as a universal language, gave rules for transforming the integers into syllables. The consonants b,c,d,f,g,h,l,m,n correspond to the digits 1 to 9, the vowels, a,e,i,o,u, to the first five powers of ten (0, 1,2,3 and 4). An integer is decomposed into a sum of powers of ten and every term of the sum is transformed into a syllable. So, the number 6, representing "man" in our example, consists of 1 term: 6×10^0 which transforms into HA. The number 65'134, for instance, would be read "HUGOBIDEFA". Leibniz mentions as an advantage of these transforming rules that the order of the syllables can be changed which should give many possibilities to poets, BIDEFA-GOHU or FABIGODEHU, and so on (Leibniz in Couturat, 1961 a).

The mapping of concepts into numbers, described above, is not only able to generate new concepts in a systematic way. It also allows to decide about the truth of propositions. For instance, an universal affirmative proposition, all A are B, is true if the number representing the concept "A" contains the number representing "B" as a factor. Of course there are many problems with this system. For instance, it would be hard to represent a disjunctive concept. Leibniz was aware of these difficulties.

Considering these few samples of the idea of Leibniz, we can imagine what enormous difficulties he was facing in trying to realize his plan. He was aware of the consequences his general logic would have, if it worked. For instance, there would be no arguing left among people, he says, because it could be decided by calculation who is right. (1880, vol.VII, e.g.p 25 f,64 f, 200). The method would be a substitute for thinking, and could be executed by a machine (Couturat, 1961 b). Leibniz was also among the first to design calculating machines.

The idea, that an algorithm could be found for every class of mathematical

problems, was apparently confirmed again and again by the further development of mathematics. However, it was finally disproved in this century for the first time by Gödel's paper, "Ueber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme", (1931). Gödel showed, that there are certain mathematical problems that cannot be solved by any algorithm out of an exactly defined class of algorithms. At this point an exact definition of the intuitive concept of algorithm had become necessary. Up to then, there had been no reason to have a precise definition. Whenever an algorithm was proposed for a certain class of problems, there was agreement whether or not it really was an algorithm. Since 1935 several precise definitions of algorithm have been given, for instance in terms of general recursive functions by Herbrand, Gödel und Kleene, in terms of Turing-machine computability by Turing, or in terms of Markovian algorithms by Markov. There is general agreement that these definitions, which have been proven equivalent to each other, capture the intuitive notion of an algorithm.

The emergence of computers has made the aspect of efficiency of an algorithm more important. Although this notion refers to all computing resources needed to execute an algorithm, one normally considers the fastest among different algorithms for the same class of problems as the most efficient one. The time required to execute an algorithm for a particular instance of a class of problems can be described as some function of the "size" or "complexity" of the instance. Such a function is called time complexity function. Different algorithms have a variety of different time complexity functions. With respect to efficiency, a simple distinction is made by computing scientists between polynomial time algorithms and exponential time algorithms. For the first group, execution time increases as a polynomial function of the size of the problem instances, and for the second group, execution time increases as an exponential function.

The practical meaning of this distinction is that problem instances with large "size" can usually not be handled by exponential time algorithms. To show this drastically, we take an example from Garey & Johnson (1979, p.7). Assume two time complexity functions, a polynomial one: n^5 , and an exponential one: 3^n ; n is a measure for the size of the problem instance. Assume

that the execution time for $n = 10$ is 0.1 sec in the polynomial case. Then it is 0.059 sec for the exponential case. Now, for $n = 40$, the time for the polynomial function goes up to 1.7 minutes, whereas the time for the exponential function goes up to 3855 centuries.

The automatic invention system of Lullus we have considered before is an example of an exponential time algorithm. Most of these algorithms are some variant of exhaustive search. Problems for which no polynomial time algorithms are known are referred to as intractable. Clearly, it is very useful to know whether a class of problems is in principle intractable, that means no polynomial time algorithm exists. In such a case, the class of problems has to be tackled by heuristic methods. In the last decade, it has become possible to prove for some classes of problems that they are decidable but that there exist no polynomial time algorithms. (Garey & Johnson, 1979).

After this rather lengthy excursion into algorithms let us go back to heuristic methods with the specification that they do not guarantee solutions. Here the work of the 19th-century mathematician Bernhard Bolzano should be mentioned who dedicated about 250 pages of his "Wissenschaftslehre" (i.e. Philosophy of Science) (1930) to heuristics. The rules which are described and commented in great detail are grouped into two classes: general rules which can be applied to any problem and special rules which are concerned with specific problems such as the analysis of concepts or propositions, or the causal analysis of events. Space forbids us mentioning more than a few of the general rules. Many of them are psychological rules such as the following ones: before one starts looking for a solution one should get a very clear picture of the problem. One should write down everything relevant to the problem in order to overcome the difficulty that the human mind can direct its attention to only a few ideas at the same time. One should avoid strong stimulation of the senses while solving a problem, and so on. Several rules concern the choice of appropriate symbols to represent the problem. Further, we find again the classification of problems into proofs and discoveries, and detailed description of the methods of analysis and synthesis. Here, Bolzano recommends combining the two methods, that means, one should work from both ends, from the data and from the hypothetical solution, at the same time.

This should help to direct the line of reasoning.

In another system, the exhaustive aspect of the older methods is taken up again. It is the morphological method of discovery developed by the astrophysicist Fritz Zwicky (1957, 1966) in this century. The first step of his method is to define the problem as exactly as possible, the second step to determine all factors influencing the solution, the third step to combine these factors and to test *each* combination for its usefulness as a solution and the final step is to choose the optimal solution. Such an exhaustive method has the important advantage that even those possibilities are evaluated which usually are missed because of various kinds of psychological blocks. Zwicky's ideas could be summarized in the following *heuristic* rule: Do *not use heuristic* rules to guide you through the solution space, but go systematically over each point of the whole space and test it carefully! Of course this procedure is only reasonable if the number of relevant factors can be kept small. Therefore, the choice of an optimal problem representation is the crucial step in Zwicky's method as his examples show.

We want to conclude our general historic review on heuristics with just mentioning the extensive and influential work of Georg Polya (1945, 1954, 1962, 1965; for a review see also Newell, 1983; van der Waerden, 1983). Polya gave detailed descriptions of the various heuristic methods and condensed them to a few general principles. By introducing brilliant examples he successfully applied them to the teaching of mathematics. The mixture of pedagogical and mathematical ingenuity made his book "How to Solve It" a great public success and contributed to the revival of heuristics as a discipline.

2. Psychology of Thinking and Heuristics

It is reasonable to assume that a psychological theory of thinking can contribute a lot to a theory of heuristics. The study of human thinking reveals specific shortcomings and limitations of problem solving and can therefore stimulate attempts to overcome them.

If we look at the early history of the psychology of thinking we notice great

differences between different schools. There even was no agreement on what the basic units of thinking were. A mode of representation had to be found or defined, before one could reasonably start talking about problem solving or creativity. We therefore will first briefly recapitulate the attempts of looking for suitable representations.

In the pre-experimental tradition up to the end of the 19th century, psychology was treated as a subdiscipline of philosophy. Since Aristotle association theory was the dominating theory of thinking. The elements of thinking were assumed to be conscious ideas or mental images. The British empirists emphasized the sensory origin of these elements. The combination of these elements was governed by one principle, the association. However, at the beginning of this century this doctrine was challenged by several schools of psychology.

First there was the Würzburg school, a group of psychologists around Oswald Külpe at the university of Würzburg, Bavaria (see Boring, 1950, or Humphrey, 1951). They made the first systematic attempt to study thought process experimentally using introspective self-reports. To their surprise they found hardly any evidence for percept-like images or clearly conscious ideas. Consequently, the Würzburg school emphasized that the elements of thought processes had neither clearly conscious nor image-like qualities, and Mayer & Orth (1901) proposed to call them 'Bewusstseinslagen' ('states of consciousness'). These 'Bewusstseinslagen' were negatively defined as being neither an image nor a percept, not an act of will and as not being further reducible. Among the other members of that group (Ach, Bühler, Marbe, Messer and Watt) it was Karl Bühler (1907) who outlined in most detail the essential aspects of a theory of thinking. He used the word 'Gedanke' to denote the fundamental units of thinking and he conceived understanding as the process of 'logical-place-assignment' of different 'Gedanken'. Although his work contains a multitude of subtle observations, his theory suffered from the lack of a more rigorous formalism for expressing his ideas. However, Bühler had anticipated many ideas of cognitive organization which became popular more than fifty years later.

Taking up the work of the Würzburg group, Otto Selz developed a more elaborate theory of thinking (For a brief outline of his arguments and his theory

see Selz, 1924). In a vigorous attack against associationism in its latest form, the constellation theory of associations by Georg Elias Müller (1911), he demonstrated that such a theory could not even account for such simple tasks as finding subordinate or superordinate concepts. Selz replaced associations having no qualities by various different links representing *directed relations*. This notion, as well as others like 'operation' or 'schema' are already very close to parallel concepts in Cognitive Science. Selz developed his own theory of productive and reproductive thinking and applied it later to the problem of improving intelligence (Selz, 1935), an issue which is taken up by de Groot (1983).

Apart from the Würzburg group and Otto Selz there was a third group working on new ideas about the structure and direction of thinking, the Gestalt psychologists. Instead of atomistic concepts, like association or 'Bewusstseinslagen', they proposed a larger basic unit of organization, the Gestalt, which is defined as a whole and cannot be reduced to a sum of its parts. Within the school of Gestalt psychologists there were three main proponents investigating thinking: Wolfgang Köhler, Max Wertheimer and Karl Duncker. Köhler (1921) and Wertheimer (1945) emphasized the central role of insight which they conceived as a sudden reorganization of a problem.

But it was probably Karl Duncker who has the strongest influence on the further developments in the psychology of thinking. First he banished the method of recording introspections used by the Würzburg group where the subjects gave an a posteriori interpretation of their inner experiences. Instead, Duncker (1935) introduced the 'solving aloud' method where subjects verbalize immediately whatever occurs to them *during* the process of thinking.

Duncker was also among the first to use a tree to represent the different intermediate steps of the solution process. This kind of representation became very popular later although some details were clearly different. In Duncker's problem solving trees successive branches represent states of increasing specificity, whereas in later applications they usually refer to successive applications of operators. This modern view of problem representation can be summarized as follows: In a first stage of problem solving, the subject has

to find out what is given, what the problem is and which actions, transformations or operators can be applied. These operators are capable of changing the states of the problem. A problem state is defined as the initial problem situation or any state that can be achieved through the successive application of allowed operators. The set of all possible sequences of operators can be represented by a solution tree, where the branches represent the application of an operator and the nodes represent problem states. A solution is defined by certain properties and all states of the solution tree which have these properties are solution states. Finding a solution then corresponds to a search for a solution state in the solution tree. Before continuing along this line, we first return to the heuristic methods proposed by Duncker.

According to Duncker, the process of thinking is essentially the resolution of a conflict between the given data and the goal. The resolution usually cannot be reached in a single step but must be achieved through a series of progressive reorganizations which are controlled by heuristic methods: "Im echten Denkprozess spielen eine entscheidende Rolle gewisse heuristische 'Methoden', die das Entstehen der sukzessiven Lösungsphasen vermitteln" (Duncker, 1935, p.28). Some of the methods Duncker proposed are (Duncker, 1935, p. 13-45) :

- Using suggestions from the given material for the solution
- Learning from mistakes through correcting variation
- 'Resonance of signal elements', i.e. using partial analogies by activating some resonance in memory
- Analysis of the conflict between the starting situation and the goal
- Analysis of the material
- Analysis of the requirements of the goal
- Availability ("Disponibilität") of the aspects of the situation, i.e. avoiding fixations on some aspects of the problem by paying attention to others
- Specialization, i.e. constructing some special cases in order to obtain a better picture of the underlying general principle
- Analysis of prerequisites.

As might be noticed, several of these methods appeared later again in the works of other authors, as, for example, Polya (1945), DeBono (1968), Adams (1974), or Wickelgren (1974), to mention just a few.

According to Duncker's view the most essential property of the available material is its *functional value* : how can something be used to obtain the goal.

At about the same time, there were other approaches emphasizing the *temporal course* of the problem solving process by postulating a series of successive stages. Here problem solving is conceived as the process of generating and testing hypotheses. The central concept is that of a *hypothesis* defined as the tentative anticipation of a result (for a review see Groner & Groner, 1982). For instance, Dewey (1933, p. 107 - 115) distinguished the following five phases of the problem solving processes: (1) suggestion, (2) intellectualization, (3) the guiding idea, hypothesis, (4) reasoning , (5) testing the hypothesis by action. In a similar proposal Claparède (1933) introduced the following consecutive 'moments': "la question, point de départ de l'opération intellectuelle; la recherche, ou découverte de l'hypothèse, ..., la vérification, ou contrôle de l'hypothèse" (1933, p.6). Still another well-known distinction was introduced by Wallas (1926): (1) preparation, (2) incubation, (3) illumination and (4) verification.

Claparède who also used solving-aloud protocols could hardly find any verbal reports about the details of the proposed procedures. Apparently there were parts of the processes inaccessible to the consciousness. This view was supported by the observations of famous mathematicians (Poincaré, 1913; Hadamard, 1945; van der Waerden, 1983) that the decisive idea for a solution very often appeared like a sudden stroke of illumination. It is therefore interesting to note that just one major heuristic rule emanated from all these approaches: wait and see, let incubation work!

As already mentioned, behaviorism - the dominating doctrine in American psychology around the forties - was not suited for application in heuristics. In the fifties several newly arising approaches had their influence on the development of psychology. We just mention some of them (for a more detailed review see Newell & Simon, 1972, p. 878ff, Newell, 1980a) : there were the new mathematical theories of information, games and decisions, and signal detection, and, of course, there were the first digital computers. These develop-

ments offered new tools for a more concise and less speculative treatment of the old questions on how thinking might proceed.

The new paradigm of great potential generality was that of information processing. Information was conceived as being invariant over different codes and independent of the particular medium or hardware. Already at an early stage, programs were developed such as the Logic Theorist (Newell, Shaw & Simon, 1959) and the General Problem Solver (Newell & Simon, 1961) or the Elementary Perceiver and Memorizer (Feigenbaum, 1959) which were conceived as computer simulations of human cognitive processes. A new discipline emerged from the joint efforts of specialists from various fields, called Artificial Intelligence.

3. Heuristics in Artificial Intelligence

Artificial Intelligence comprises all efforts at making computers solve problems that require intelligence such as, for example, playing chess, proving logical theorems, making medical diagnoses, designing and analyzing electronic circuits, understanding language or images etc. Artificial Intelligence can be seen from two different viewpoints, from an engineering point of view and from a 'scientists' point of view. From an engineering point of view, its goal is to produce machines that are increasingly powerful in increasingly diverse areas of application. From the second, more general point of view, Artificial Intelligence is a new approach to determine the prerequisites of intelligent behavior and to uncover its nature, independent of whether this behavior is produced by a natural, biological or an artificial system. In the last few years this second view has become very popular and has led to a new scientific paradigm called 'Cognitive Science' which pursues this goal in an interdisciplinary effort between Artificial Intelligence, cognitive psychology, linguistics and other disciplines. In Cognitive Science both man and machine are uniformly seen as systems that construct symbolic descriptions of the world and manipulate these descriptions. One consequence of this view is the separation of the physical realization of the symbols from their meaning and thus it leads to the view that the study of the computation processes on the symbols

can be distinguished from the study of the physical mechanisms that support the computation (For a more detailed discussion of these issues see Newell, 1980b). From this position it is then natural that the results found in Artificial Intelligence can be carried over to man and can be used to understand his own intelligent behavior.

The history of Artificial Intelligence is intimately connected with the notion of heuristics, or to be more precise, the notion of heuristic search. Two important ideas lie behind this notion. First it is the idea that intelligent behavior can be characterized as search, the search for solutions to problems posed to the system. Secondly it is the idea that the search space associated with problems requiring intelligence is typically so large that any system must use heuristics to reduce it to a manageable size given the system's limited resources of time and space. 'Time' here refers to the number of operations the system must perform to find a solution while 'space' refers to the number and size of symbolic descriptions of intermediate solutions the system produces on the way to a solution. Lullus' automatic invention system has lead us to consider algorithms with exponential time complexity and we have seen that such algorithms exceed any practical time limits for large problem instances. We could avoid this problem by using a parallel search scheme but this would lead to an explosion of intermediate solutions produced by the system and thus would violate the constraint of limited space resource.

Heuristics are used to reduce the search space of a given problem. But what are heuristics? The ideal answer would be a formalized notion of heuristics and we might gain as much out of such a formalization as we gained from the formalized notion of an algorithm. Unfortunately things look not that promising: The history of Artificial Intelligence shows us that heuristics are difficult to delineate in a clearcut manner and that the convergence of ideas about their nature is very slow.

This is not the place to go into a detailed discussion about the role of heuristics in Artificial Intelligence., but a few very basic remarks seem appropriate to judge the contributions of Artificial Intelligence towards a theory of heuristics.

Early work in Artificial Intelligence has led to a set of techniques on how to use general knowledge to guide the process of generating new steps in the search space. These techniques, known as heuristic search techniques, grew out of a large number of empirical studies and are described, for example, in Nilsson (1971). But these studies have also shown that the use of very general knowledge alone is usually not sufficient to constrain the search space down to a manageable size. As a heuristic rule one can say that any piece of knowledge, as specific it may be, should and in practice must be used to constrain the search space. With the increased size of knowledge bases we are faced with a whole bunch of difficult problems: How must all this knowledge be represented in a system? How should knowledge be used, i.e. how can a piece of knowledge lead to appropriate actions by the system? How should knowledge be organized, i.e. how do we manage that a system has access to all knowledge relevant to a situation but is not diverted too much by knowledge irrelevant to that situation? How do we manage that any piece of knowledge including knowledge about partial solutions is available to all those components of the system that may usefully apply it? How can a system acquire the huge mass of knowledge it may need, i.e. how can we transfer our knowledge to the machine or how can the system discover that knowledge by itself?

A particularly promising aspect of Artificial Intelligence is the fact that it provides us with tools for empirically studying our ideas on a future theory of heuristics (e.g. Lenat, 1983).

4. Common Properties of Heuristic Methods

At this point we should briefly summarize the common aspects of heuristic methods as they are used today :

- (1) Problems (and their solution) are *represented in term of information processing* (computer programs, terms of automata theory, cognitive models).
- (2) Heuristic methods *restrict the search space* to a manageable size.
- (3) They are *sufficient* with respect to a certain output, thus generating a sequence of operations from the problem to the solution. With ill defined or open-goal problems this property cannot be assessed in advance.

- (4) They are expected to provide a better solution than a pure trial and error procedure (*efficiency*).

This list is not complete, and still idealized since, as has been seen in earlier chapters and will also be seen below, several approaches called "heuristic" do not satisfy all of these properties. We are still far off from an adequate definition and comprehensive theory of heuristics as it exists for algorithms¹⁾ One consequence of this fact is that it is impossible to generalize from a particular instance of a heuristic to classes of problem domains, unless on a purely speculative basis. It is also impossible to derive something like an analytic solution over a whole parameter space, as e.g. in statistical estimation. Therefore the efficiency characteristics of a heuristic have to be explored empirically either by comparing systematically the predicted performance of different heuristics with each other (Groner, 1978; Groner & Groner, 1982) or by simulation studies (Lenat, 1983).

5. Heuristics in Judgment and Decision

Since there is no universally accepted definition for decision (neither in current research nor in the present book) which would clearly rule out neighbouring areas like judgment, inference, etc., we will also include examples from these areas in this chapter, being well aware that our presentation is highly selective and not even representative for all fields of decision research.

It is commonly accepted that heuristic approaches in decision have been made popular by Herbert A. Simon (1955, 1956) with the two related concepts of "bounded rationality" and "satisficing". In the conventional additive utility models of decision research it had been assumed that the homo oeconomicus would behave rationally by optimizing his expected utility. The amount of

1) In a conference on methods of heuristics (University of Bern, September 1981; proceedings published in Groner, Groner & Bischof, 1983) one of the discussion themes was the question "How could heuristics be defined?" The comment of Marvin Minsky to this question was "Give me a definition of life". Needless to say that neither of the concepts were adequately defined at that conference!

computation involved in arriving at an optimal decision was not taken into account, an algorithmic decision scheme was implicitly assumed.

Simon argues against this conceptualization: "The point of departure is the observation that human thinking powers are very modest when compared with the complexities of the environments in which human beings live. If computational powers were unlimited, a person would simply consult his or her preferences (utility functions) and choose the course of action that would yield maximum utility under the given circumstances. That is, of course, just what the "rational man" of classical economic theory does. But real human beings, of bounded rationality, cannot follow this procedure. Faced with complexity and uncertainty, lacking the wits to optimize, they must be content to satisfice - to find "good enough" solutions to their problems and "good enough" courses of action.

The first specification, then for a model of thinking processes is that it be a model of bounded rationality, that it incorporate mechanisms for coping with complexity even when it is unable to digest it completely ... Satisficing provides an escape from the difficulty that, in a complex world, the alternatives of action are not given but must be sought out. Since the search generally takes place in a space that is essentially infinite, some stop rule must be imposed to terminate problem solving activity. The satisficing criterion provides that stop rule: search ends when a good-enough alternative is found"(1979, p.3).

The work on judgmental biases by Tversky & Kahneman bears some relation to the notion of bounded rationality. "When faced with the difficult task of judging probability or frequency, people employ a limited number of heuristics which reduce these judgments to simpler ones" (Tversky & Kahneman , 1973, p. 207).

By the representativeness heuristic, "people predict the outcome that appears most representative of the evidence. Consequently, intuitive predictions are insensitive to the reliability of the evidence or to the prior probability of the outcome, in violation of the logic of statistical prediction" (Kahneman

& Tversky, 1973, p.237). However, representativeness as an explanatory concept suffers from a serious difficulty: it is much easier to state *a posteriori* that some evidence must have been representative for a certain outcome, than to make a *prediction* for which outcome an evidence will be representative.

The availability heuristic suffers from similar difficulties: "A person is said to employ the availability heuristic whenever he estimates frequency or possibility by the ease with which instances or associations could be brought to mind" (p. 208). In some sense, availability can be interpreted as an attentional factor which regulates the search process either in memory or in a perceptual domain. This leads immediately to the same objection as with the representativeness heuristic, i.e. circularity of argumentation: if we do *not* have direct access to the attentional process but only infer from the subject's response what he ought to have attended to, the model will become unfalsifiable. The amount to which attentional processes can be manipulated experimentally is limited. According to Tversky & Kahneman (1973) the empirical evidence for the availability heuristic is the positive correlation between a frequency estimation of some event and the number of instances produced by the same individual. We might polemically object that Tversky & Kahneman fell into the trap of their own heuristic "causal scheme" by taking these correlations as evidence for causation.

With respect to our list of common properties of heuristic methods outlined in chapter 4, we also notice that "heuristics" in the sense of Tversky & Kahneman lack almost all important attributes since they are not stated in terms of an unambiguous language such as information processing theory. This makes them vulnerable to criticisms like the ones stated above.

Recently there is a strong trend towards models of decision based on the information processing approach (e.g. Payne, 1976; Hogarth, 1980; Huber, 1982; for reviews see Einhorn & Hogarth, 1981). The process as a whole is decomposed in a series of subprocesses, like information acquisition, evaluation, action, and feedback/learning (Einhorn & Hogarth, 1981) which interact with each other and are built up of still smaller subprocesses. Three different

sources of data are used as empirical evidence for the postulated processes: information pick-up from display boards (e.g. Payne, 1976), registration of eye fixations (Russo & Rosen, 1975; Groner, 1978; Groner & Groner, 1983), and solving-aloud protocols (Einhorn, Kleinmuntz & Kleinmuntz, 1979; Huber, 1982).

Although the vocabulary of most approaches seems to display much communality on a superficial level, a closer examination gives a more variegated picture. Apparently it is difficult to put the different research pieces together, since - as we argued above - we still lack even the outline of a comprehensive theory of heuristics. It can only be hoped that all the present prolific model building and assiduous experimentation will lead us somewhat closer to that goal.

References

- Adams, J.L.: 1974. Conceptual blockbusting. San Francisco: Freeman
- Amarel, S.: 1938. Problems of representation in heuristic problem solving: Related issues in the development of expert systems. In: R. Groner/ M. Groner & W.F. Bischof (Eds.), Methods of heuristics. Hillsdale, N.J.: Lawrence Erlbaum
- Bolzano, B.: 1930. Wissenschaftslehre, Vol. III. Leipzig: Meiner
- Boring, E.W.: 1950. History of experimental psychology. New York: Appleton - Century-Crofts
- Bühler, K.: 1907. Ueber Tatsachen und Probleme zu einer Psychologie der Denkvorgänge. I. Ueber Gedanken. In: Archiv für die gesamte Psychologie 9, 297-365
- Claparède, E.: 1933. La genèse de l'hypothèse. In: Archives de Psychologie 24, 2-155
- Couturat, L.: 1961a. Opuscles et fragments inédits de Leibniz. Hildesheim: Olms
- Couturat, L.: 1961b. La logique de Leibniz. Hildesheim: Olms
- DeBono, E.: 1968. New think. New York: Basic Book
- Descartes, R.: 1908. Oeuvres, Vol. X. Adam, C. & Tannery, P. (Eds.), Paris: Léopold Cerf

- Dewey, J.: 1933. *How we think*. Second edition. Boston: Heath
- Düncker, K.: 1935. Zur Psychologie des produktiven Denkens. Berlin: Springer.
(English translation: Duncker, K.: On problem solving. In: *Psychological Monographs* 58, 1-113)
- Einhorn, H.J. & Hogarth, R.M.: 1981. Behavioral decision theory: Processes of judgment and choice. In: *Annual Review of Psychology* 32, 53-88
- Einhorn, H.J./Kleinmütz, D.N. & Kleinmütz, B.: 1979. Linear regression and process-tracing models of judgment. In: *Psychological Review* 86, 465-485
- Feigenbaum, E.: 1959. An information processing theory of verbal learning. Progress Report P-1817, RAND Corporation
- Garey, M.R. & Johnson, D.S.: 1979. *Computers and intractability*. San Francisco: Freeman
- Gödel, K.: 1931. Ueber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. In: *Monatshefte für Mathematik und Physik* 38, 173-198
- Groner, R.: 1978. *Hypothesen im Denkprozess*. Bern: Huber
- Groner, R. & Groner, M.: 1982. Towards a general hypothetico-deductive theory of cognitive activity. In: R. Groner & P. Fraisse (Eds.), *Cognition and eye movements*. Amsterdam: North-Holland
- Groner, R. & Groner, M.: 1983. A stochastic hypothesis testing model for multiterm series problems based on eye fixations. In: R. Groner/D.A. Fisher & R.A. Monty (Eds.), *Eye movements and psychological functions: International views*. Hillsdale, N.J.: Lawrence Erlbaum
- Groner, M./Groner, R. & Bischof, W.F.: 1983. Approaches to heuristics: A historical review. In: R. Groner/M. Groner & W.F. Bischof (Eds.), *Methods of heuristics*. Hillsdale, N.J.: Lawrence Erlbaum
- de Groot, A.: 1983. Heuristics, mental programs, and intelligence. In: R. Groner/M. Groner & W.F. Bischof (Eds.), *Methods of heuristics*. Hillsdale, N.J.: Lawrence Erlbaum
- Hadamard, J.: 1945. *The psychology of invention in the mathematical field*. Princeton, N.J.: Princeton University Press
- Hankel, H.: 1874. *Zur Geschichte der Mathematik in Altertum und Mittelalter*. Leipzig: Teubner
- Hogarth, R.M.: 1980. *Judgment and choice: The psychology of decision*. Chichester, U.K.: Wiley
- Huber, O.: 1982. *Entscheiden als Problemlösen*. Bern: Huber
- Humphrey, G.: 1951. *Thinking. An introduction to its experimental psychology*. London: Methuen, and New York: Wiley

- Kahneman, D. & Tversky, A.: 1973. On the psychology of prediction. In: *Psychological Review* 80, 251-273
- Kircher, A.: 1669. *Ars magna sciendi sive combinatoria*. Amsterdam: Jansson
- Köhler, W.: 1921. *Intelligenzprüfungen an Menschenaffen*. Berlin: Springer
- Leibniz, G.W.: 1880. *Philosophische Schriften*, Vol. I-VII. In: C.J. Gerhardt (Ed.), Berlin: Weidmannsche Buchhandlung
- Lenat, D.B.: 1983. Toward a theory of heuristic. In: R. Groner/M. Groner & W.F. Bischof (Eds.), *Methods of heuristics*. Hillsdale, N.J.: Lawrence Erlbaum
- Lullus, R.: 1656. *Ars Magna*. Frankfurt: Cornelius Sutorius
- Mayer, A. & Orth, J.: 1901. Zur qualitativen Untersuchung der Association. In: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 26, 1-13
- Müller, G.E.: 1911. Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufs. Leipzig: Barth
- Newell, A.: 1980a. Reasoning, problem solving and decision processes: the problem space as a fundamental category. In: R.S. Nickerson (Ed.), *Attention and Performance VIII*. Hillsdale, N.J.: Erlbaum
- Newell, A.: 1980b. Physical symbol systems. In: *Cognitive Science* 4, 135-184
- Newell, A.: 1983. The heuristic of Georg Polya and its relation to Artificial Intelligence. In: R. Groner/M. Groner & W.F. Bischof (Eds.), *Methods of heuristics*. Hillsdale, N.J.: Lawrence Erlbaum
- Newell, A./Shaw, J.C. & Simon, H.A.: 1958. Elements of a theory of human problem solving. In: *Psychological Review* 65, 151-166
- Newell, A. & Simon, H.A.: 1961. GPS, a program that simulates human thought. In: H. Billing (Ed.), *Lernende Automaten*. München: Oldenbourg (reprinted A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill 1963)
- Newell, A. & Simon, H.A.: 1972. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall
- Nilsson, N.J.: 1971. *Problem-solving methods in Artificial Intelligence*. New York: McGraw-Hill
- Payne, J.W.: 1976. Task complexity and contingent processing in decision making: an information search and protocol analysis. In: *Organizational Behavior and Human Performance* 16, 366-387
- Poincaré, H.: 1913. *Mathematical creation*. In: G.B. Halsted (Ed.), *The foundation of science*. New York: Science Press
- Polya, G.: 1945. *How to solve it*. Princeton, N.J.: Princeton University Press

- Polya, G.: 1954. Mathematics and plausible reasoning, Vol. I + II. Princeton, N.J.: Princeton University Press
- Polya, G.: Mathematical discovery. New York: Wiley. Vol. 1: 1962; Vol. 2: 1965
- Russo, J.E. & Rosen, L.D.: 1975. An eye fixation analysis of multialternative choice. In: Memory & Cognition 3, 267-276
- Selz, O.: 1935. Versuch der Hebung des Intelligenzniveaus. Ein Beitrag zur Theorie der Intelligenz und ihrer erziehlischen Beeinflussung. In: Zeitschrift für Psychologie 134, 236-301
- Simon, H.A.: 1955. A behavioral model of rational choice. In: Quarterly Journal of Economics 69, 99-118
- Simon, H.A.: 1956. Rational choice and the structure of the environment. In: Psychological Review 63, 129-138
- Simon, H.A.: 1979. Models of thought. New Haven & London: Yale University Press
- Tversky, A. & Kahneman, D.: 1973. Availability: A heuristic for judging frequency and probability. In: Cognitive Psychology 5, 207-232
- van der Waerden, B.L.: 1983. Inspiration and thinking in mathematics. In: R. Groner/M. Groner & W.F. Bischof (Eds.), Methods of heuristics. Hillsdale, N.J.: Lawrence Erlbaum
- Wallas, G.: 1926. The art of thought. New York: Harcourt & Brace
- Wertheimer, M.: 1945. Productive thinking. New York: Harper
- Wickelgren, W.A.: 1974. How to solve problems: Elements of a theory of problems and problem solving. San Francisco: Freeman
- Zwicky, F.: 1975. Morphological astronomy. Berlin: Springer
- Zwicky, F.: 1966. Entdecken, Erfinden, Forschen im Morphologischen Weltbild. München: Droemer Knaur

ON THE TESTABILITY OF THE AVAILABILITY HEURISTIC

KLAUS FIEDLER

*Universität Giessen
Fachbereich Psychologie
F.R. Germany*

In this article I am concerned with the theoretical idea that social judgments are mediated or even determined by the *availability* of relevant information, that is, by the ease with which different pieces of memorized information come to the judge's mind (Tversky & Kahneman, 1973). I will mainly focus on one empirical phenomenon which has been explained in terms of an availability heuristic, namely, the egocentric bias in responsibility attributions, according to Ross & Sicoly (1979) or Thompson & Kelley (1981). When the partners in a close personal relationship judge their own and their partner's contribution to various activities (such as planning leisure activities or breaking promises), they tend to overestimate their own contribution. This egocentric bias has been explained by reference to the empirical fact that Ss are more likely to recall examples of their own activities than of their partner's activities. Hence, differential recall products are assumed to cause the judgment bias. I am primarily concerned with this causal assumption.

As Ross and Sicoly (p. 334) correctly admit, there is "no definitive evidence that the bias in availability *causes* the bias in responsibility." Indeed, the main evidence is merely correlational and it may be noted, parenthetically, that this is true of most applications of the availability concept. For example, in Thompson and Kelley's Study 2, the mean correlation between egocentric judgments and the tendency to recall self-related instances was $r=.50$. But such a piece of correlational evidence may be due to a spurious rather than a causal relation. Thus, if an S's judgment that he/she was more responsible for breaking promises coincides with recalling that S himself/herself broke a promise, this may simply reflect the fact that he/she

has really broken more promises than his/her partner. To the extent that recall and judgments are veridical, they will tend to be correlated even though the process of judgment formation may be independent of any retrieval processes.

Indeed, the theoretical weakness of explaining judgmental outcomes in terms of the availability heuristic originates in the very difficulty of distinguishing operationally between recall measures and judgmental measures. Memory-based judgments can also be regarded as recall of relevant information, and vice versa. Thus, the two operations involved -- i.e., judging that one has broken more promises and recalling that one has broken more promises -- are similar to such a degree that any causal explanation of the former in terms of the latter must appear theoretically empty.

Trivially, any memory-based judgment involves recall operations and activates relevant memory content. However, the non-trivial version of the availability principle, as I understand it, states that judgmental deviations reflect recall biases operating at the time of judgment formation. This is essentially an assumption about the *process* of judgment formation, emphasizing the role of retrieval attempts *within* that process.

The present research constitutes an attempt to overcome the methodological difficulties inherent in testing the causal role of differential recall in judgment formation. Egocentric biases were investigated by having the participants perform two cognitive operations: (a) a judgment operation, that is having Ss judge the percentage of their own (vs. their partner's) contribution to 20 different activities; and (b) an elementary recall operation, that is, having them recall *one* example for each activity. Judgment of an activity item, say breaking promises, either preceded or followed the recall of an example of that item. Thus, the temporal position of the two tasks was varied.

The main dependent measures were the response latencies, in particular, the facilitation effect (i.e., latency reduction) of prior recall on subsequent judgments and of prior judgment on subsequent recall. Of course, facilitation can be expected in both directions. As recall and judgment operations for the same activity refer to the same knowledge base in memory, there should be a priming effect of the first operation and a consequent reduction in latency for the second operation. However, an availability-based explanation has implications for the *relative degree of facilitation* in the two directions.

According to the differential recall explanation, the recall operation can be regarded as a subroutine, as it were, of the judgment task; that is, a minimal recall operation is included as part of the judgment process. For the arguments that follow, it is sufficient to state this inclusive relation more loosely as in the left part of Figure 1. The graphical representation means that if recall is a subprocess of judgment formation, more elements of an elementary recall task must be included in the judgment task than vice versa. Thus, if the judgment operation is performed prior to recall, there should be most facilitation, for most elements of a recall operation should be already activated as part of the preceding judgment task. On the other hand, if recall is performed first, there should be less facilitation in later judgments, for a relatively smaller part of the judgment operation is acti-

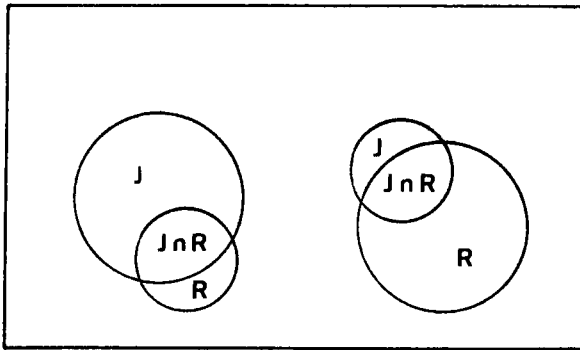


Figure 1. Graphical representation of two possible relations between recall and judgment operations. (J and R are the sets of cognitive elements making up judgment and recall, respectively.)

uated by the preceding recall task. Therefore, the idea that recall of relevant information is a mediating subprocess in the process of judgment formation entails more facilitation in the judgment-first condition.

An alternative theory -- that recall is the more inclusive operation -- is depicted in the right part of Figure 1. Of course, the predictions are opposite in this case.

An implicit assumption underlying this kind of reasoning ought to be

made explicit. The assumption states that facilitation is a positive monotonic function of the degree to which the to-be-primed material is included in the primary task (i.e., the degree of overlap), and that response time is independent, within certain limits, of other features of the priming task (i.e., the non-overlapping part). Although there are still some problems, this assumption is at least consistent with common priming manipulations (cf. Wyer & Srull, 1981; Higgins & King, 1981).

Now let us consider the experimental design. Two independent variables were crossed. The first factor refers to the comparison of the judgment-first and the recall-first condition, as already mentioned. The other factor refers to the manipulation depicted in Figure 2. The two tasks of making judgments and recalling instances were either performed *blockwise* or in an *alternating* fashion. In the blockwise condition, the first task (e.g., judgment) was first completed for all items before turning to the second task (e.g., recall) with the same 20 items in the same order. In the alternating condition, however, both tasks were completed for the same item before proceeding to the next item of the series, thus switching between the different tasks from trial to trial.

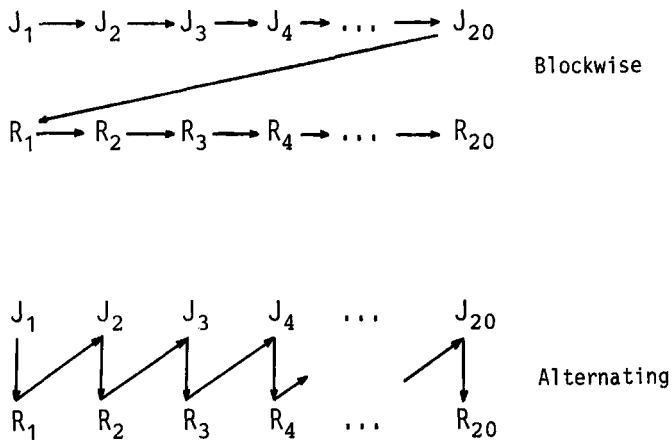


Figure 2. Schematic representation of sequences of trials in the blockwise and alternating condition.

Clearly, the facilitation effects should be more pronounced in the alternating condition where the two operations on the same item are performed in close temporal succession and are not separated by much distracting activity, as in the blocked condition. However, it might be that facilitation in the alternating condition is so pronounced that differences with respect to recall or judgment-first cannot be sensibly detected. Hence, the blocked condition, where much less facilitation can be expected, might afford a more sensible test.

Before I come to the empirical results, let me briefly summarize the main predictions to be tested:

- The phenomenon to be explained must be replicable, that is, egocentric biases should occur and there should be a correlated bias to recall self-related examples.
- According to the usual availability-based explanation, facilitation in the judgment-first condition ought to be more pronounced.
- Further, there should be substantial correlations, within judges across the 20 items, between recall latencies and judgment latencies, if recall is the determining subprocess in judgment formation.
- And these correlations should be higher in the alternating condition, due to the temporal closeness.

The material, consisting of a subset of 20 activities selected from Thompson and Kelley's 36 activities, appears in Table 1 along with the mean percentage judgments of Ss' own contribution. Obviously, the egocentric bias was replicated, that is, the participants judged their own contribution to be more than 50%, and this tendency is highly significant ($t=5.48$; $p<.001$). Moreover, there is a correlated tendency to recall self-related examples. Or stated another way, judgments of Ss' own contribution increase for those items in which a self-related example is recalled ($t=2.86$, $p<.02$) and decrease for those items in which a partner-related example is recalled ($t=3.98$, $p<.001$). Having replicated the explanandum, we can now turn to the latency data.

Response latencies were transformed to yield normalized distributions and then transformed into standard z-scores. Then the mean latency *per item* was calculated separately for each experimental condition. The resulting 20 blocks of two by two latency values of each item were subjected to analyses of variance.

There were two reliable effects in the recall as well as the judgment data (Table 2): Highly significant main effects for task order and substantial

Table 1
 Social activities used as stimulus material
 along with mean judgment of Ss' contribution
 (Items are retranslated from German into
 English)

Planning joint leisure activities	52.4 %
Irritating the other	56.8 %
Carrying the conversation when the two of you are alone	58.8 %
Criticizing the other	57.9 %
Demonstrating affection for partner	50.9 %
Complaining about the other	54.5 %
Resolving conflicts between the two of you	54.7 %
Causing arguments that occur between the two of you	57.2 %
Being sensitive to the other's needs	54.9 %
Acting possessively toward the other	51.2 %
Expressing fears and worries	55.4 %
Considering partner's interests when planning activities	50.8 %
Maintaining self-control during an argument	46.3 %
Helping partner with his/her problems with people and work	50.6 %
Embarrassing partner in public	59.6 %
Initiating discussions about the relationship	56.8 %
Being assertive in your relationship	54.8 %
Making an effort to spend time with partner	56.8 %
Accepting criticism from partner	46.5 %
Taking out on partner one's outside frustrations	51.4 %

interactions. As Table 2 shows, the time required to judge is much shorter if relevant information has been recalled before, that is, in the recall first condition (the main effect). However, the interaction indicates that this is largely due to the alternating condition. Nevertheless, the small facilitation effect of the blocked condition is still significant at the 5%-level.

Turning now to the recall latencies, there is again much facilitation in the alternating condition, that is, reduced latencies for recall in the judgment-first condition. However, this time there is no facilitation whatsoever in the blocked condition.

Table 2
Mean statistics and ANOVA results for judgment latencies and recall latencies. (Values in parentheses refer to untransformed data)

Judgment latencies:		Judgment first	Recall first	Significant effects
Blockwise		0.20 (5.47)	-0.29 (4.04)	Main effect columns: $F_{1,19}=56.99$; $p<.001$
Alternating		1.01 (7.82)	-0.92 (3.31)	Interaction: $F_{1,19}=14.99$; $p<.002$
Recall latencies:		Judgment first	Recall first	Significant effects
Blockwise		0.16 (12.80)	0.06 (12.84)	Main effect columns: $F_{1,19}=68.63$; $p<.001$
Alternating		-1.18 (7.65)	0.96 (16.04)	Interaction: $F_{1,19}=83.47$; $p<.001$

This pattern of results is invariant with regard to the data transformations. Table 2 also includes the untransformed values which yield essentially the same picture: Reduction of recall *and* judgment latencies by more than 50% in the alternating condition; and in the blocked condition, no recall facilitation but considerable judgment facilitation. Hence, a task situation has been identified where there is facilitation in only one direction. Contrary to an availability-based explanation, only prior recall facilitates later judgments, suggesting that an elementary recall task is the more inclusive operation. The recall task is more time-consuming (as the absolute latencies reveal) and is experienced as requiring more effort. Even minimal recall (recalling but one example) seems to be a deeper cognitive process than judging one's contribution on a percent scale.

If the discrepancy between recall facilitation and judgment facilitation (in the blocked task situation) were only a matter of degree, it would be more difficult to interpret because we do not know which time unit of recall laten-

cy corresponds to which time unit of judgment latency. Fortunately, however, the discrepancy is a qualitative one in that the facilitation effect is confined to only one direction.

Further negative evidence against the assumption that recall mediates or even determines the process of judgment formation stems from correlational analyses. The correlations between recall latencies and judgmental latencies (computed within judges across the 20 items) are remarkably low and are not higher in the alternating as compared to the blockwise condition ($t=0.21$; $\bar{r}_{alt} = .20$, $\bar{r}_{block} = .18$). If recall operations were the most essential sub-routines of the judgmental process, we would have expected somewhat higher correlations.

The reader may further have recognized that in both latency ANOVAs there was no main effect for the alternating versus blocked comparison. A final glance at Table 2 clarifies why this is the case: In the alternating condition, the gain in response time for the task in the *second* position is cancelled out by an *increase* of response time for the *first* task. For example, the recall latencies of about 17 sec for the recall-first condition are well above the 12 sec for the blocked conditions which may be regarded as a baseline (the same inhibition appears with respect to judgment latencies). Obviously, switching between recall and judgment tasks (in the alternating condition) interferes with the cognitive processing. This subsidiary finding is also hardly compatible with the assumption that recall is a naturally occurring subprocess embedded in the process of judgment formation.

Let us finally consider the content of the recall outputs. As already noted, the correlation between egocentric attributions and a tendency to recall self-related examples should not be interpreted to mean that recall is causally involved in the process of judgment formation. Pooling across all judges and items, 19% of the recall responses are self-related, 7% are partner-related, and 74% have no clear person reference. Further, 41% of the same recall responses can be regarded as concrete episodic events, 52% as (proto-) typical examples without episodic reference, and 7% are not codable in this respect. Hence, the most typical recall output is uninformative with regard to which partner contributed most in the past. It seems, therefore, that the cognitive conditions for an availability-based mechanism are simply not fulfilled.

Concluding Remarks

On the whole, the evidence reported here is clearly inconsistent with Ross and Sicoly's (1979) and Thompson and Kelley's (1981) availability account for the phenomenon of egocentric attributions. The nature of the facilitation effects, the within-judge correlations between recall latencies and judgment latencies, the contents of the recall outputs, and the inhibition effects in the alternating condition all seem to indicate that the judgmental process is not causally dependent on recall operations. One might argue that the present operationalization is inappropriate: Implicit recall operations that are involved in the process of judgment formation cannot be experimentally simulated by such an explicit recall task. I am principally inclined to agree with this argument. However, it should be noted that the positive evidence which seems to support the availability conception rests on the same questionable operationalizations. Of course, a research method that is acceptable for confirmation must also be acceptable for disconfirmation.

There is no reason to generalize the present findings to all kinds of social judgments and to completely dismiss the availability theory. We can imagine many situations in which judges have to rely on whatever happens to be available in memory, especially when required to make retrospective judgments on past issues that they have never thought of before. In many other cases, however, the social judge can already anticipate the requirement to judge at the time when information is perceived and encoded in memory. It is most likely in these cases that judgments are already formed at the early encoding stage and incorporated as an integral part of the memory representation (cf. Lingle & Ostrom, 1979). Intuitively, I would suggest that most judgments made in this experiment about Ss' own and their partner's contributions belong to this class of pre-formed judgments.

Incidentally, it should be noted how willingly and prematurely the availability hypothesis has been accepted and used to explain other judgmental phenomena. It seems curious to me that the idea has never been really challenged. As I understand the nature of scientific growth, strong and fruitful theories always give rise to theoretical and empirical controversies. But there is virtually no controversy around the availability theory (or some version of it).

Consider, as another example, the availability-based account of illusory correlations (Chapman & Chapman, 1967, 1969), that is, the illusion that

a correlation could be observed in a series of events where actually none was present. Since Tversky & Kahneman's (1973) suggestion that "the various sources of illusory correlations can all be explained by the operation of a single mechanism -- the assessment of availability" (p. 224), repeated attempts have been made to explain illusory correlations via selective recall but, to my knowledge, not a single published experiment has tried to demonstrate that the origin of the illusion may be already located in the perception or encoding of the observations. In my own (still unpublished) research on the judgment of covariation, I have repeatedly found that judgmental deviations can be (partly) predicted from individual judges' subjective interpretations of the stimulus events (i.e., from encoding). Moreover, using a free recall test in five experiments -- as a measure of which stimulus observations are spontaneously available in memory -- I was never able to support a differential recall mechanism. Hence, my own evidence does not fit very well with the picture that emerges from the literature. Here I have recognized again that a lack of controversy -- a biased search in favor of an availability explanation -- characterizes the research on illusory correlations.

My intentions here, and my conclusions, can be summarized as follows:

First, I wanted to direct some attention to the fact that certain research tools that are common for the cognitive psychologist -- such as analyzing response latencies -- have not been utilized to substantiate the notion of judgmental heuristics.

Second, there are too few overt controversies in the field of person memory and judgment, in the light of the serious problems a critical-minded researcher must realize in existing theories. Controversies reflect the growth of scientific knowledge.

Third and finally, the monistic search for availability phenomena, that is, for retrieval effects on judgments, may have distracted researchers from an equally promising approach, namely, to investigate the role of encoding effects in judgment formation.

References

- Chapman, L.J. & Chapman, J.P. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, 72, 193-204.

- Chapman, L.J. & Chapman, J.P. Illusory correlations as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, 74, 271-280.
- Higgins, E.T. & King, G. Accessibility of social constructs: Information-processing consequences of individual and contextual variability. In N. Cantor & J.F. Kihlstrom (Eds.) *Personality, cognition, and social interaction*. Hillsdale, N.J.: Erlbaum, 1981.
- Lingle, J.H. & Ostrom, T.M. Retrieval selectivity in memory-based impression judgments. *Journal of Personality and Social Psychology*, 1979, 37, 180-194.
- Ross, M. & Sicoly, F. Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 1979, 37, 322-336.
- Thompson, S.C. & Kelley, H.H. Judgments of responsibility for activities in close relationships. *Journal of Personality and Social Psychology*, 1981, 41, 469-477.
- Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, 5, 207-232.
- Wyer, R.S., Jr., & Srull, T.K. Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E.T. Higgins, C.P. Herman, & M.P. Zanna (Eds.) *Social cognition: The Ontario Symposium (Vol. 1)*. Hillsdale, N.J.: Erlbaum, 1981.

This Page Intentionally Left Blank

PROBLEM STRUCTURING CALCULI AND LEVELS OF KNOWLEDGE REPRESENTATION IN DECISION MAKING

PATRICK HUMPHREYS

DINA BERKELEY

*London School of Economics and Political Science
Decision Analysis Unit
and Social Psychology Department
U.K.*

Introduction: Difficulties with the representation of knowledge in decision making tasks

The thesis we present in this paper is that in any study of individual decision making it is essential to take into account, quite explicitly, the social context in which the decision making activities take place, and within which the decision maker conceptualizes the consequences of his or her actions.

This thesis is not new, in fact it was the foundation for Bernouilli's (1738) discussion of how a 'prudent' person should choose between alternative acts, in terms of the impact that the possible consequences would have on the immediate social context; if the prudent person was a merchant, this would be his business, and so on. However, these ideas formed part of normative decision theory, prescribing the form of 'economic rationality' that should characterize a prudent businessman, or whatever. Analysis of how the social context is taken into account in the process of decision making has largely been concerned with examining how actions chosen and inferences made by people involved in decision making tasks compare with the prescriptions of the representation of the decision making task in a normative model of the situation.

Proceeding in this manner has effectively excluded from analysis the two aspects of decision making with which this paper will be principally concerned. These are, (i) the way that knowledge which the decision maker believes to be relevant is represented, and, (ii) how, on the basis of this knowledge, the decision maker generates a structure representing his or her view of the problem, through a linkage between key elements (acts, events, consequences) and within which content (beliefs, values, preferences) can be explored and manipulated in searching for a prescription for action.

The reason why these aspects have been excluded is that in conventional decision making experiments, the representation of the decision making task arrives **pre-structured** at the start of the experiment. Within what Kahneman and Tversky (1982) called the 'conversational paradigm', the experimenter has in mind a representation of the task which he or she conveys to the subject through the experimental instructions. All the knowledge required for constructing a 'normatively correct' structure is presumed to be contained in these instructions, although in some experimental situations it is assumed quite likely that the subject will ignore part of this information (e.g. information about base rates in a reference population) and develop an **incomplete** task structure (c.f. Bar-Hillel, 1982).

In 'real life' decision making, everything is not so neat. Instructions are not neatly available at the outset of each decision task, specifying the relevant knowledge and indicating how to structure it. Often structures do not arrive ready for use, but have to be **generated** in representing decision problems, so that preferences and uncertainties can be explored in deciding what to do. In this wider context it is our thesis that the social world in its many manifestations as culture and language in the macrolevel and as family, school, work environment, peer group, etc., in the microlevel not only determines the terrain of definitions and the range of understanding an individual member can use and perform in his or her decision making activities but also provides the person with the arena in which his or her choices and actions will be tried out, justified, interpreted by others and rewarded or sanctioned. Under such an understanding of the relationship of the individual to his or her environment, the individual is seen not as passive, acted upon by the environment, but as active, interpreting, acting upon and changing that environment.

Adopting this view has implications for the types of analytic models that may be used in characterizing decision making. Under the 'conversational paradigm', with its use in context-isolated mainly laboratory based tasks, it seemed quite reasonable to take over models developed by psychologists in perceptual judgement tasks, and much of the terminology used in psychological decision making models, like 'anchoring and adjustment' (Tversky and Kahneman, 1974) or 'veridical perception' (Tversky and Kahneman, 1981), points to these roots (c.f. Johnson, 1972).

With this view of judgement came the emphasis on 'characteristics of the receiver' of the information supplied (c.f. Swets, Tanner and Birdsall, 1961) in terms of looking for suboptimal performance (according to the normative standard) or 'bias'. This emphasis was pervasive; the notion of 'heuristics' was originally associated with innovative activity with the aim of increasing scope (Newell and Simon, 1972) or efficiency (Thorngate, 1980) in problem solving activities, saving energy and time. However, in accounts of decision making experiments within the conversational paradigm, it became linked with the notion of **bias** to the extent that the two terms seem to have acquired nearly equivalent meaning (c.f. Nisbett and Ross, 1980; Kahneman, Slovic and Tversky, 1982). In later papers following this approach (e.g., Kahneman and Tversky, 1979) one can find whole networks of heuristics being pieced together to account for complex patterns in biases which appear when the results of a number of experiments are considered together.

All this assumes that the instructions presented to subjects provide a repository of knowledge concerning the task situation which is sufficient to permit the construction of a problem representation within which a subject's actual decision making performance can be compared with that which is considered to be 'normatively correct'. However, in cases where we have analysed the way subjects actually structure decision problems on the basis of instructions prescribed for typical tasks within the conversational paradigm (e.g. those cited by Tversky and Kahneman 1974, 1981; Kahneman and Tversky, 1979) very large individual differences have been found in the way subjects structure the tasks (Phillips, 1983). Moreover, these differences are not due to varying degrees of 'incompleteness' when compared with the structure cited as being normatively correct by the writer of the instructions. Instead, they result from elaborations generated by the

subject when faced with under-determined and incomplete problem descriptions (the usual case). The resulting problem representations are not necessarily 'sub-optimal' compared to those used by the experimenter in computing the 'normatively expected' inferential values. Indeed the opposite is usually the case: subjects were often found to create better elaborated structures providing a more complete description of the actual task context within which the subjects located themselves.

We do not wish to imply the conclusion that "experimenters should try to write better instructions for their subjects". Rather, we would like to stress the impossibility of ensuring that subjects with a variety of different goals and different histories of decision making activities will ever generate fully equivalent problem representations for any particular set of instructions given to them in a laboratory setting. It is even less likely that such communality could ever be ensured in any real-life context, in which each decision maker will bring his or her different history, and different referents for the meaning of consequences, etc. (Berkeley, 1982). In our opinion, the crucial point to note in studying the effects of heuristics and biases influencing human decision making is that the inferences made and values reported by subjects are conditional not only on the information processing heuristics employed but also, and more importantly, on the form of the particular structure generated by the subject as his or her view of the representation of the problem. Hence improving decision making, or training decision makers, requires first and foremost that we understand and aid the structuring process.

Conditional rationality: The importance of 'fixing' structure

Decision theory has traditionally been silent about issues involved in problem structuring. Part of this reason has been the central position of the idea of 'economic rationality' in the theory, and the guiding role it has played since the beginning. When discussed as a concept within the literature, rationality is usually taken to refer to the axioms of the theory, like coherence, transitivity, and adherence to the subjective

expected utility principle. This definition of the concept we shall call the **theoretical definition** in the sense that it articulates those which, in theory, are the characteristics of a rational actor or action.

However, testing the existence of these principles or lack thereof in any given actor is achieved only via an operationalization of this definition. The **operational definition** of rationality involves the employment of formal models whose construction is founded on these axioms. When operationalized, these formal models essentially consist of composition rules (Krantz et al, 1971) operating within an axiomatised decomposition of the phenomena of interest, which in decision theory are immediate courses of action.

For example, SEU provides an additive composition rule prescribed when immediate acts are decomposed through act-event sequences into consequences at the decision horizon which, in turn, are assigned utilities (Krantz, 1972). MAUT provides a composition rule for obtaining values of these utilities when they themselves have been decomposed into relative preference values on attributes characterizing those consequences (Humphreys, 1977); and so on.

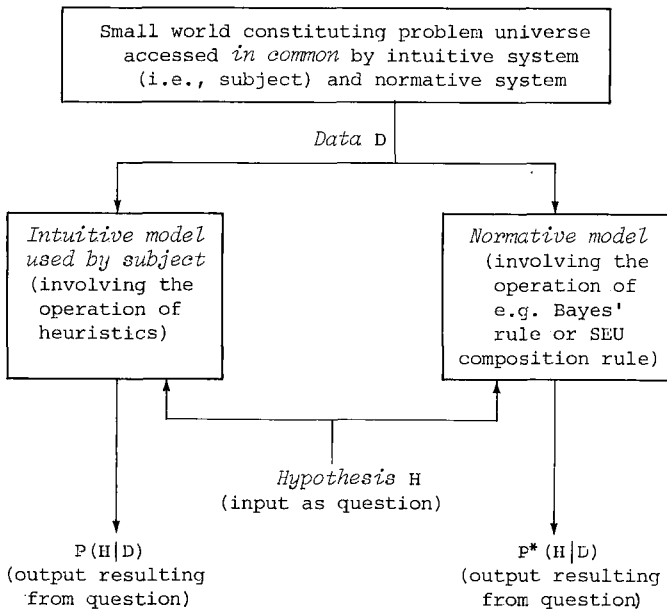
Typically, empirical testing of a decision theoretic composition rule involves a comparison procedure like that shown in figure 1. In cases where a discrepancy is noted between the two outputs in figure 1, the usual argument is either to criticize the adequacy of the model, (and maybe discussing how to replace it with one whose output converges more closely with that obtained from subjects, c.f., Kahneman and Tversky's 1979 "Prospect Theory") or taking the model as specifying the ideal type, and identifying the differences in outputs as a 'bias' on the part of the subject. 'Corrective action' would then take on the characteristics of 'debiasing' (Fischhoff, 1982).

In either case, while empirical testing is carried out on the operational definition of the concept of rationality, the results of the testing are taken to reflect rationality or irrationality at the theoretical level. However, such inferences only have meaning if a common understanding exists between subject and experimenter about ends, means and representation of knowledge in the problem structure within which the comparison of data

indicated in Figure 1 is to take place (Hovland, 1952).

In other words, comparisons like those indicated in figure 1 police **conditional rationality**. That is, rationality can only be investigated conditional on achieving a common understanding of the **form** of the problem structure within which **content** is manipulated, whether according to the subjects' intuitive calculus, or according to a theoretic formal calculus axiomatized on the basis of 'rational' assumptions.

Figure 1: Comparison procedure used in generating data for empirical testing of a decision theoretic composition rule



(source: Berkeley and Humphreys 1982, Figure 2)

This common understanding serves to temporarily 'fix' the form of the decision problem representation within which inferences will be drawn and preferences revealed by all the parties involved (subject and experimenter;

opposing parties in treaty negotiations, interest groups in social decision making, etc.). Berkeley and Humphreys (1982) discuss how 'fixing of the form' of a decision problem may involve the resolution of no less than seven different types of uncertainty.

Two of these concerned first, uncertainty about how the decision maker will feel, and wish to act, having arrived at a subsequent act choice point after intervening events have unfolded in real time (Elster, 1979) ; and second, uncertainty about the extent to which one possesses agency for inducing changes in the probabilities of subsequent events through being able to alter relations between states of the world (Savage, 1954; Vari and Vecsenyi, 1983). Handling these two of the seven types of uncertainty at an individual level involves resolving existential problems concerning the position of one's self or the other in the social context and lie outside the scope of this paper (Parkes, 1971, discusses issues of this type in terms of 'psychosocial transitions').

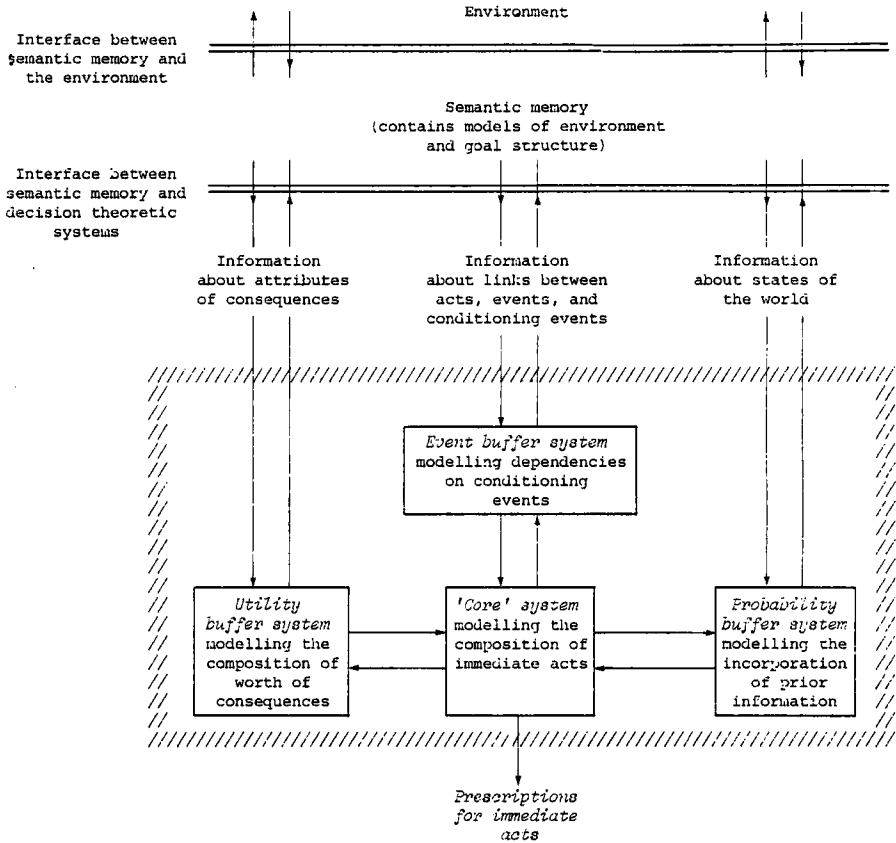
Four of the remaining types of uncertainty can be taken into account in representation of decision problems, given the current state of decision theory. These relate to (i) the probabilities of outcomes of subsequent events, conditional on what has preceded them in the act-event sequence , (ii) the probabilities of outcomes of subsequent events conditional on other extraneous events , (iii) how to incorporate prior information, and (iv) how to conceptualize the worth of consequences and, when more than one criterion of 'worth' is involved, how to combine these criteria.

Berkeley and Humphreys (1982) discuss how decision theory provides the basis for representing the way in which these four types of uncertainty were resolved in representing any particular decision problem through four different **types** of structure, each handling one type of uncertainty. These are shown in the boxes in figure 2, together with the interfaces between them which may be involved in developing a complete problem representation.

Figure 2 represents the current state of development of the 'subjective' view of decision theory pioneered by Edwards (1954; Edwards, Lindman and Phillips, 1965). All these systems are considered to be interfaced directly with the decision maker's semantic memory. Decision problem representations built

Figure 2: Systems formalized within Decision Theory

The four systems (core system, event buffer system, utility buffer system and probability buffer system) are of different structural types, but the interfaces between them are fully specified within the decision theoretic framework.



(source: Berkeley and Humphreys 1982, Figure 1)

through the use of these systems are constructed on the basis of knowledge represented in memory but do not in themselves describe knowledge structures within memory. Neither, it should be pointed out, do they describe objects in the physical world. 'States of the world' may be named at nodes located within decision theoretic structures, but what is represented is the decision maker's belief concerning the subjective probabilities of these states of the

world's occurrence under some scenario of the future, mediated through his or her knowledge structures in semantic memory.

This raises the fundamental problem of how to resolve the seventh, and perhaps most fundamental type of uncertainty, which Hogarth, Michaud and Mery (1980) called **procedural uncertainty**, and described as "uncertainty concerning means to handle or process the decision, e.g. specifying relevant uncertainties, what information to seek, and where, how to invent alternatives and assess consequences, etc." Decision theory, in its conventional form, is quite silent about how to handle this type of uncertainty. Yet handling, and resolving **procedural** material is always involved in translating the generalized possibilities for decision theoretic structures shown in figure 2 into any actual problem representation which could be considered to be 'requisite' (Phillips, 1982) in representing and exploring a particular problem. Any attempt to formalize ways in which procedural uncertainty may be resolved implies the development of a problem structuring calculus.

Attempts to build a problem structuring calculus

Berkeley and Humphreys (1982) identified three different approaches to building a problem structuring calculus. The first involves constructing a simple taxonomy of tasks where analytic characteristics of alternative decision theoretic structures are matched to task characteristics (Brown and Ulvila, 1977, Howell and Burnett, 1978). Two difficulties are involved here. First, the same problem can be structured within any one of a number of quite different analytic frameworks (Brown, 1978) with no normative way of deciding between them. Secondly, according to von Winterfeldt (1980), "substantive rather than analytic characteristics identify real problems".

Von Winterfeldt (1980) reviewed and rejected existing taxonomies, which, following the first approach, used mainly analytical categories to distinguish problems as being ill-suited to lead a decision maker to a class of problems which had similar characteristics for the problem under consideration. Instead, he followed the second approach to building a problem structuring calculus : developing a problem taxonomy, based on

substantive rather than analytic features of tasks. This taxonomy was designed to be problem-driven, but still generalizable. For example, one could find a 'prototypical structure' for budget allocation in one cell of the taxonomy, a 'prototypical structure' for facility siting in another cell, and so on.

The idea behind storing prototypical structures in the taxonomy is that these would still be generalizable. In any specific budget allocation problem, one would elaborate the appropriate prototypical structure supplied by the taxonomy to meet specific task requirements. The problem is that even when only prototypical structures are defined, building up a taxonomy of them according to substantive features of problems still requires a great number of cells, each of which has to be investigated on an ad-hoc basis.

A third approach, which we said we preferred, was to employ a calculus comprising more powerful operators than the simple 'search and match' required in operating a taxonomy. We suggested that it should be possible to develop a **generative problem structuring calculus** using production system rules (c.f. Anderson, 1976, sect.3.3) operating in a way analogous to the re-writing rules proposed by Chomsky (1957) as the basis for generative grammars: 're-writing' structural elements into the complete structure within which content can then be manipulated. This idea was proposed by Bonczek, Holsapple and Whinston (1981), as a method for generating decision support systems, but they limited their discussion to systems where the structures were computer programs, and re-writing was performed on the computer language 'structural elements' (i.e. statements within the program). To us, it seemed that this idea might well have a much wider application. Chomsky (1957) and Bonczek et al (1981) had independently considered two specific, but quite different sets of grammatical rules (those of English and those of Fortran). Fillmore (1968), however, has discussed how more general grammars can be constructed for use in substantive applications. The key property of Fillmore's 'case grammar' approach is that in such a grammar, structural elements are 'tagged' according to the **functional** relation each element provides (the 'case' it takes) when linked into the structure. Hence, case grammars, unlike Chomsky's purely syntactic, transformational-generative grammar, are problem driven and with this comes all the advantages which von Winterfeldt (1980) pointed out.

However, this still leaves one very crucial problem: how should one specify the nature of the structural elements to be manipulated as content by case-grammar type production systems in generating decision problem structures over the enormous range of applications and contexts where decision making activities are carried out? If a single-level structure-generating scheme were proposed, the number of elements would be vast, as general-purpose 'natural' languages have to rely on very large vocabularies.

Thus, it would be futile to attempt to specify all the structural elements in a calculus generating decision problem representations at a single level. As an alternative this paper investigates what a **multi-level** problem structuring calculus might comprise, and how this leads to multiple levels of knowledge representation concerning decision making tasks. The problem-structuring systems we reviewed above already comprise two levels: a lower level calculus manipulating content within defined structure (c.f. figure 1) and one operating at a higher level of abstraction in generating that structure within a particular (assumed temporarily fixed) problem context.

But why stop at just two levels of abstraction? To do this would be to overlook the insight in Piaget's (1972, 1977) account of genetic epistemology that, in accounting for the progress of knowledge, there is no reason to stop at the second level of abstraction. In a 'two level' decision making scheme what is represented as **form** at the first level, is manipulated as **content** at the second level. But this relation may be continued through further levels of abstraction, that is, what is represented as form at the second level, can be manipulated as content at the third level, and so on. This implies, of course, that the content manipulated at each level is qualitatively different, as are the progressively more powerful operators which become available to a person who can understand the principles underlying their employment within the calculus formalized at each progressive level of abstraction.

Adopting such an approach greatly simplifies the process of decision problem modeling : a problem may be represented at a number of levels of conceptualization, but we may be able to provide a powerful and relatively simple description of the problem at each of the levels involved. The success of such an enterprise, however, hinges on the fact that we are

proposing a multi-level scheme for conceptualizing decision problems. Where is the evidence that decision makers themselves conceptualize problems at a number of levels, and how are these conceptualizations related to the context of the decision problem?

The most comprehensive review of evidence for these propositions that we know of is that provided by the studies of bureaucratic organizations carried out by Elliott Jaques and his colleagues (Jaques 1976, 1982; Jaques et al, 1978). While Jaques' account of 'levels of abstraction' within the hierarchy of roles occupied by personnel in such organizations does not constitute a microcosm which can serve as a model of any and all social contexts, it has the advantage that it allows us to examine the relationship between levels of responsibility, levels of demand characteristics of tasks and levels of knowledge representation in structuring the decision problems involved in these tasks. In other words, it provides a conceptual laboratory for exploring possible characteristics of a multi-level scheme in a variety of practical settings in the way we illustrate in the next section. The paper concludes by suggesting some implications for the study of decision making in general, beyond the confines of activities carried out within idealized bureaucracies.

Levels of abstraction in decision problem representations

The left hand column in table 1 gives a brief summary of the equivalence established by Jaques (1976, 1982) between levels of abstraction of role, responsibility (expressed in terms of 'time-span') and of demand characteristics of decision makers within bureaucratic organizations.

According to Jaques, the qualitative differences between the levels of organizational role shown in the first column can be understood in terms of progressive levels of abstraction in the symbolic construction of actions which may be carried out by executives at each level. Moreover, these levels are not viewed as a specific product of organizational forms, rather bureaucratic levels are parasitic on levels of abstraction of (idealized) tasks within organizations.

Table 1: Comparison of demand characteristics of tasks facing personnel having responsibilities at a given organisational level with structuring compatibilities required in representing decision problems at that level.

Level	Organisational level in bureaucratic employment hierarchy	Time span inherent in problem representation at given level	Demand characteristics of tasks facing personnel with responsibility at given level	Structuring capabilities required in representing decision problems at given level
5	Corporate subsidiary/ enterprise (managing director work system)	5 - 10 years	Problems not dealt with in context set wholly from above can modify boundaries of business within policy or i.e., define work system.	Articulation of principles guiding conditional (goal) closing of an open system, and/or re-opening of a conditionally closed system.
4	General management (of e.g., development production or sales and within work system.	2 - 5 years	Detachment from specific cases, seeing them representative examples of issues calling for development of a system.	Selecting/interfaces capability between structural types (DSS generation from scratch or DSS generator construction).
3	Department managerial/ principal specialist	1 - 2 years	Control of trend of tasks and problems arising. Extrapolation from trend to ways of formulating problems.	Re-structuring capability within single fixed structural type (e.g., criteria generation in multi-criteria model).
2	Front-line managerial/ professional	3 months to 1 year	Formal operational, can anticipate changes in tasks due to any one of: demand, object, production resource, pathway, or pathway resource.	Manipulation of data on one variable at a time within fixed structure (e.g., sensitivity analysis).
1	Shop and office floor	less than 3 months	Concrete operational, limited to tasks concretely and physically at hand.	Estimation of values at nodes within fixed structure (e.g., informational retrieval system).

The second column in table 1 gives the typical time-span in the problem representation which is requisite for a task carried out by an executive at each level. Time span indicates how far away the decision horizon (Toda,1976) is set for tasks for which the decision maker is held responsible within the organizational context.

For example, a person whose role in an organization is set at level 1 might be given the responsibility involved in minding a machine tool. It would matter for his career prospects within the organization if he failed to decide on appropriate action given concrete evidence of impending malfunction of that machine. This same person may also gain evidence of organizational malfunction at a higher level (e.g., through noticing that the manufactured product is beginning to look increasingly outmoded, compared with those of competitors which he or she has looked at in high street shop windows). In this case he or she may describe to colleagues what "should be done about this", or even submit this information for consideration at a higher level through a "suggestion box". However, he or she would not be expected to **decide** on the proposed action, or be held responsible for its consequences.

The third column in table 1 summarizes the description of the demand characteristics of the tasks facing personnel with responsibility at a given level in an organization given by Jaques (1982, pp 80-85). Jaques was primarily concerned with bureaucratic organizations comprising idealized hierarchies of work strata, and in any actual organizational context we may find personnel at particular organizational levels also responsible for carrying out tasks at lower levels (rather than delegating them to subordinates) or taking initiatives at more than one level where the role structure permits (e.g., as in the case of consultants or "problem fixers"). However, the precision of the equivalence between organizational roles and task characteristics is not our major concern here. Instead, we were particularly interested in the way Jaques described the **qualitative** differences between tasks at the various levels of abstraction ('qualitative' differences are always implicit in the use of the term 'level'; see Bunge, 1969).

It suggested to us that we might find parallel levels of abstraction in structuring capabilities required in generating representations of decision

problems. The results of our preliminary investigations in this regard are outlined in column 4 of table 1, and are discussed with some brief examples below.

Level 1

Starting at the bottom of table 1, we find problem processing being handled at the concrete operational level (level 1). Piaget (1972) describes key principles which have to be understood in controlling operations at this level as those involved in **seriation** ($A < B$; $B < C$, etc.) and the inter-connected postulates of **transitivity** and **conservation**. Note that it is the understanding of precisely these principles which allows one to adhere to what Edwards, Lindman and Phillips (1965, pp 272-275) described as four relatively uncontroversial principles of rationality commonly cited in decision-theoretic accounts of coherent efficient decision making, namely decidability, transitivity, dominance and the sure-thing principle.

This implies that achievement of the concrete operational level within the Piagetian account of the development of thought is a prerequisite for being able to handle decision problems in a coherent way (c.f. Brown and Lindley, 1982). However the operations actualized at this level are limited to providing best assessments of quantities (e.g., position on a probability or on a preference scale) to be represented as content at a defined node in a fixed problem structure. In other words, the type of question addressed here is **what is the best assessment** of a probability of a particular event in the current context? A good example of decision making activities located at this level is Zimmer's (1983; this volume) account of processing of subjective probability assessments (from the subject's rather than the experimenter's viewpoint). Efforts to improve the quality of decision making located exclusively at this level have centered around attempts to ensure that decision makers are "well-calibrated" (Lichtenstein, Fischhoff and Phillips, 1982). In each case the problem is seen as how to ensure that the answer given represents the decision maker's **best assessments** based on his or her current knowledge of states of the world.

Level 2

Moving up to level 2 involves understanding the principles underlying what Piaget (1972) called **formal operational thought**. Piaget describes the fundamental characteristic of cognitive abilities at this level as the "capacity to deal with hypotheses instead of simply objects", hypotheses which are expressed as propositions rather than facts.

The **qualitative difference** between adjacent levels is always that what is **represented as form** at the lower level is **manipulated as content** at the higher level. Piaget (1971; 1972) describes the key formal operations at level 2 as comprising the **INRC group**: operationalizing principles relating to Inversion, Negation, Reciprocity and Correlation. While each of these types of operations may be involved in a decision maker's attempt to provide 'best assessments' in level 1 problem solving, it is at level 2 that these properties **as a group** are first understood and exploited in exploring aspects of a (pre-structured) representation of a decision problem.

In text books on decision analysis (e.g. Brown, Kahr and Peterson, 1974) this type of exploration of aspects of a problem is generally referred to as "sensitivity analysis": exploring "what if?" questions about changing values at nodes in the structure (e.g., the probability of an event in an act-event sequence) to see what effects are propagated throughout the structure. Phillips (1982) describes a case study showing the importance of such activities in "requisite" decision modeling in enabling an industrial company's board to agree on the decisions involved in the development and launching of a new product.

Hence, at level 2 the content manipulated within the structure is not "facts" but hypotheses (opinions, views, etc.). It is explicitly recognized that probabilities can vary according to different scenarios of future states of the world and that utilities vary in reflecting the range of specific participants' interests and preferences in group decision making. A decision maker's knowledge base supporting his or her problem processing at this level has to comprise procedures supporting the manipulation of content which can be matched to specific "what if" enquiries.

The possibilities for re-structuring a decision problem representation at level 2 are nevertheless extremely limited. Jaques'(1982) account of demand characteristics of problems which can be handled by personnel working at this level stresses that they are limited to those problems where changes in circumstances will require that changes in only **one** major parameter need be investigated at a time (see table 1). When changes in only one are investigated, the rest of the problem structure remains undisturbed, acting as the frame within which these changes may be explored. In sensitivity analyses within a decision tree, for example, changes in probabilities at event nodes are "framed" by the current act-event structure connecting these nodes to consequences with utilities which remain fixed within the analysis, so that the impact of these changes on expected utility calculations can be assessed.

Level 3

More fundamental re-structuring involving changes in relations between two or more of the major parameters handled in isolation at level 2 requires problem processing at level 3 or above. At level 3, we move beyond Piaget's account of INRC operations dictating the form of the current problem representation. They now become content manipulated within operations aimed at re-structuring the problem under the constraint that the **type** of structure used to represent the problem remains the same. The meaning of 'type' here can be interpreted by reference to figure 2, above. Each box identifying a decision-theoretic subsystem in that figure marks out techniques for generating problem structures of one particular type (act-event linkage, multi-attribute utility decomposition, and so on). Re-structuring activities which are superordinate to reordering within a single structural type are handled at level 4 and above in ways we shall outline later.

For an example of decision making activities at this level, figure 3 gives an extract from a summary of an interactive session with a decision aiding system called MAUD4 (Humphreys and Wisudha, 1982), which possesses problem structuring capabilities at level 3.

In this instance, the decision maker was primarily seeking assistance in understanding an aspect of the **structure** of his problem, which was to choose

between the four job offers listed at the top of figure 3. The problem was to identify the criteria which really mattered in choosing between them.

In working with MAUD4, the decision maker identified nine criteria, of which three were retained in the final structure (the fifth, eighth and ninth in order of identification). Making a choice between alternative job offers implies for most people choosing between alternative futures affecting a large part of their lives. Consequently the number of facts which are potentially relevant in identifying criteria is very large, and so no attempt was made in this case to build a machine-based knowledge system to provide

Figure 3: Extract from the summary produced by MAUD4 of an interactive session with a person facing a job decision problem

```
SUMMARY OF SESSION DAVID EVANS SO FAR:

Current order of preference of JOBS from best to worst
(preference values are given in brackets)

COMMUNITY TV (0.78) <BEST
AMERICA FREELANCE (0.51)
WELSH ARTS CENTRE (0.32)
CRITIC (0.00) <WORST

Ratings of JOBS on the scales you are currently using

          C   C   W   A
          O   R   E   M
          M   I   L   E
          M   T   S   R
          U   I   H   I
          N   C   C   C
          I           A   A
          T           R   F
          Y           S   R
          T           E   E
          V           C   E
                   E   L
                   N   A
                   T   N
Rating scale      R   C
number           E   E

( 5 )   5   7   3   4   CULTURALLY FULFILLED(1) to CULTURALLY ISOLATED(9)
                        Ideal value= 1
( 8 )   4   6   5   3   FEELING ALIVE(1) to FEELING DEPRESSED(9)
                        Ideal value= 2
( 9 )   3   5   1   5   CREATIVE(1) to BORING(9)
                        Ideal value= 3
```

* The following information represents your preferences for the JOBS under consideration, on each of the scales. 1.00 represents the best JOB and 0.00 represents the worst JOB on each scale.

* If you wish to change anything, you may do so at the end of this summary.

ATTRIBUTES CURRENTLY IN THE PERSON'S INDIVIDUAL PREFERENCE STRUCTURE		C	C	W	A	
		O	R	E	M	
		M	I	L	E	
		M	T	S	R	
		U	I	H	I	
		N	C		C	
		I		A	A	
		T		R		
		Y		T	F	
				S	R	
		T			E	
		V		C	E	
				E	L	
Rating scale number				N	A	
				T	N	
				R	C	Rating scale name
				E	E	
{	(5)	0.50	0.00	1.00	0.75	CULTURALLY FULFILLED to CULTURALLY ISOLATED relative importance = 0.20
	(8)	0.67	0.00	0.33	1.00	FEELING ALIVE to FEELING DEPRESSED relative importance = 0.36
	(9)	1.00	0.00	0.00	0.00	CREATIVE to BORING relative importance = 0.44

The following scales are no longer in use for the reasons given below

ATTRIBUTES INTRODUCED AND LATER DELETED AS A RESULT OF LEVEL 3 RE-ORDERING ACTIVITIES	{	(1)	NATURAL to MUCH EFFORT This scale was cancelled because of its similarity with scale number 2
		(2)	CLOSE TO FAMILY to ESTRANGED FROM FAMILY This scale was cancelled because of its similarity with scale number 1
		(3)	TRUEVOCAATION to DETERMINATION AND DESPAIR This scale was cancelled because of its similarity with scale number 4
		(4)	FEELING AT HOME to FEELING ALONE This scale was cancelled because of its similarity with scale number 3
		(6)	ADVENTUROUS to INDEPENDENT MATURITY This scale was cancelled because of its similarity with scale number 7
		(7)	CHALLENGING DIFFICULTIES to ESCAPE FROM RELATIONS This scale was cancelled because of its similarity with scale number 6
			END OF SUMMARY.

the information which would identify the criteria. Instead, the knowledge system within MAUD4 organizes procedures for eliciting, investigating, and reordering criteria, that is, structural elements rather than facts about states of the world. The domain of the knowledge system is thus brought within tight bounds, but the knowledge it organizes is represented at a quite different level from the 'facts' typically organized within automated decision support systems (c.f. Bonczek, Holsapple and Whinston, 1981; McCosh and Scott Morton, 1978).

In our view, the general characteristic of level 3 problem structuring calculi is that they manipulate as content procedures which serve to define form at level 2, just as level 2 structuring calculi manipulate propositions that define form at level 1. MAUD4 represents a specific implementation of such calculus, operating at level 3 in interaction with the decision maker, and manipulating level 2 procedures like those listed above in generating a structure to represent the decision problem. However, within MAUD4 this structure is always expressed in terms of a set of attributes on which part-worths (Kneppreth et al, 1974) of the set of options currently under consideration may be represented. Thus, the constraint under which MAUD4 operates is that this is the requisite type of structure for handling the current decision problem. Here, this constraint was acceptable, as the decision maker was facing a situation where his options were already defined in terms of a set of immediate acts (acceptance of job offers), each of which was directly linked to a single consequence (consequent career) whose (complex) total worth had to be assessed relative to the others.

Field tests with various versions of MAUD (Humphreys and McFadden, 1980, Bronner and de Hoog, 1983) indicated that in cases where MAUD was successful in aiding decision making, this was due to its ability to aid the decision maker's attempts to **restructure** the problem in such a way that the goals on which the structure is predicated are clarified, so that a coherent composition rule can be applied in exploring the relative desirability of the options under consideration. In other words, the key to success lay in the adequacy of MAUD's capabilities at level 3.

In some cases, though, problem structuring constrained to level 3 will not be acceptable, as when the immediate problem is to decide on what type of structure to adopt in developing the problem representation. In deciding upon a new career, such a problem could arise when one wished not only to be able to evaluate current options (involving one type of problem representation), but also to be able to develop new options (e.g, through generating new act-event-consequence scenarios - like "form one's own company....."). Choosing the **type of problem representation** which is requisite at any point in a decision making process is in fact a level 4 problem (see below). At level 3, only the **results** of this process are evident, where they serve to define the type of problem structure within which level 3, level 2, and level 1

activities are carried out as, for example, in MAUD4.

Level 4

Level 4 is concerned with the development of decision making activities of much wider scope. Table 1 indicates that the demand characteristics of problems represented at this level involve detachment from specific cases, seeing them representative of issues calling for the development of a system. Whereas a **specific** case may be handled through specific problem representations developed at level 3, structuring capability at level 4 involves the articulation of principles which manipulate level 3 problem structuring modules as content. There is no formal language within decision theory which articulates these principles, as they are superordinate to the forms addressed by decision theory, shown in the boxes in figure 2. Hence in working at level 4 a decision maker has either to articulate these principles within his or her own natural language or to learn a new language enabling the efficient and coherent generation of systems linking level 3 (sub)problem representations into a structure specifically addressed to modeling the whole range of aspects of the problem under consideration. Hence the language employed by the decision maker serves as a generative problem structuring calculus at this level.

Figure 4 shows the results of the use of a 'language calculus' of this type by the manager of the marine claims department of a large insurance company working in interaction with a decision analyst. The aim was to generate a system which could provide for decision making concerning the expected size of a marine claim, following notification at any time in the future of an incident involving any one of a number of large ships insured with the company (this is called a "hull claim"). Details of the use of this calculus and the level 3 modules it manipulates were given in Humphreys, Wooler and Phillips (1980, section 6). Here we shall outline how it fits into our discussion of decision problem representation at level 4.

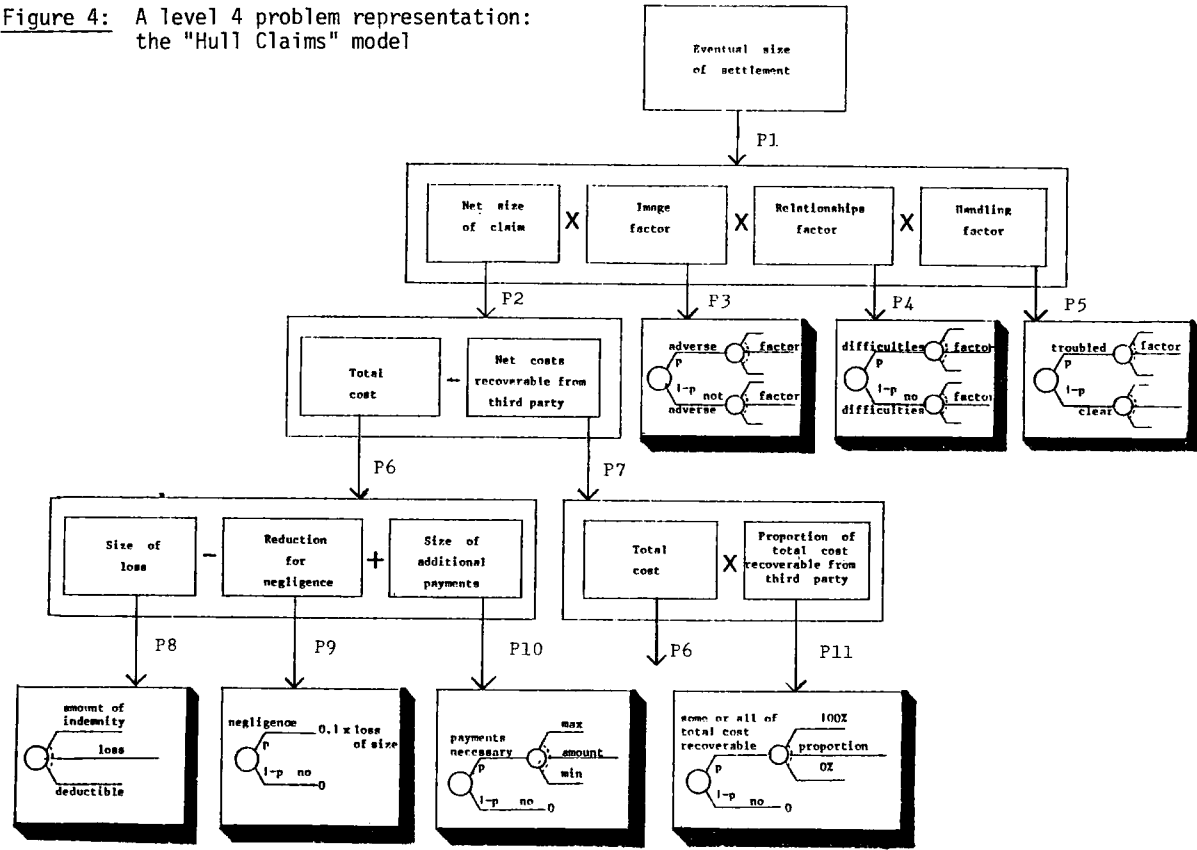
Unlike our level 3 'MAUD' example here the problem processing system was not automated: it originated and remained in the mind of the decision maker who used his own 'natural' language in developing the problem representation

shown in figure 4. His goal was to be able to assess a single complex consequence (eventual size of settlement) whatever the circumstances. From the way he discussed this consequence it was evident that it could be "re-written" in terms of the net size of any claim, inflated by the appropriate "image factor", "relationship factor" and "handling factor". These factors sound like jargon to one outside the insurance field. They do not exist **within** decision theoretic terminology, because they are names which index level 3 decision theoretic problem representations: handling how to take into account the uncertainty about (i)whether the way the claim is settled will affect the company's image adversely, (ii) whether it will lead to difficulties in relationships with other clients, and (iii)whether the process of handling the claim will be troubled or clear.

The decision analyst worked with the decision maker in developing the structure of the problem at level 4. However, the aid that the analyst could provide at this level was **not** in teaching the decision maker how to use a problem structuring calculus. In fact the opposite was the case: he had to **learn** the calculus (including the language 'jargon') which the decision maker was using intuitively. However, the analyst was able to capitalize on the relationship between re-writing rules (in this case those identified in the decision maker's use of language) and production systems in formalizing the representation of the structure the decision maker was generating, and thus aid his attempts to check its coherence.

In figure 4 you can see how the decision maker's production systems we discussed earlier, enable the decision maker's rewriting of "eventual size of settlement" to be expressed as production P1, and for the three 'factors' to be expressed as productions P3, P4 and P5, each of whose right hand side is a terminal node in the level 4 system: a level 3 module. The decision analyst was able to develop the level three models on the basis of structures described within the province of decision theory shown in figure 2. At this lower level, it was the analyst rather than the decision maker who provided the language although the decision maker was primarily interested that his subordinates in the appropriate company departments should **understand the principles** at this level : they would need to articulate these structures in any instance, and report the results to him. Support for these level 3

Figure 4: A level 4 problem representation:
the "Hull Claims" model



structuring activities is, of course, required at levels 2 and 1 in ways analogous to those we described for the level 3 system MAUD4 above.

Unlike P3, P4 and P5, the right hand side of P2 is not a terminal node in the level 4 system. When the decision maker was asked about the net size of the claim, he said it was a function of the "total cost" minus the "net costs recoverable from a third party", each of which he wished to express in his own natural language, rather than as decision theoretic modules ("total cost" is the "size of the loss" minus any "reduction that can be obtained for negligence" plus "any additional payments which have to be made to other parties"). Hence the right hand sides of P6 and P7 are also non-terminal nodes, as at these nodes the decision maker wished to continue structuring the problem at level 4.

A sceptic might ask here: why does the hull claims problem have to be expressed at level 4? In the case where all the modules are of the same structural type, it should be possible to design a system directly linking them together, and even if they are not, one could implement the interfaces shown in figure 2 between modules of different structural types, so why do we need this fancy level 4 production system?

The answer is as follows: first, while it would be quite easy to design a system in the way just described **given** the structural relations outlined in figure 5, it would not have been possible to **generate** this system from scratch without a level 4 analysis of the problem: one would be without the essential information of how the subproblem representations should be linked together. In the hull claims decision making problem, only the decision maker (the marine claims manager) **understood the principles** required in deciding how to do this, even though his understanding was only intuitive, and expressed in natural language. At level 4, the decision analyst could only help him formalize this understanding. At level 3, the decision analyst's training provided him with an intuitive understanding of how to generate event trees structuring relations between uncertain states of the world, and so he could take a much more active role in developing the structures at this level (the form of each module), and explaining their operation to the decision maker. In other words, the knowledge representations at level 4 were accessed from the decision maker's semantic memory, but those at level 3 were

accessed from the decision analyst's semantic memory. At level 2 the knowledge representations in the decision analyst's memory were accessed (he developed the sensitivity analyses used to "prune" the modules so just the key uncertainties would be represented), but at level 1, the decision maker provided the estimates of the probabilities at all the event nodes.

Thus, in generating the system from scratch, either the decision maker or the decision analyst working with him had to understand the principles underlying the operations, and provide the knowledge base at each of the four levels. However, once the system had been generated, its use in eventual size of settlements no longer required level 4 decision making. The output from the level 4 problem processing system **codifies** the context within which each level 3 module is addressed to a subproblem, and can provide the brief for a manager at level 3 in the organizational hierarchy of the company (see Table 1, column 1) on the information his department needs to provide. Similarly the event tree structures resulting from level 3 problem structuring within each module can codify the framework within which information is to be accessed (as best estimates of the probabilities which apply in a specific case) at level 1 and subjected to sensitivity analysis at level 2.

The idea behind codification in this way has a long history. Examples are the book of Leviticus in the Bible (c.f. Leach, 1964), and tactical handbooks for military field commanders. Understanding how to construct such books requires structuring knowledge at a relatively high level, which then provides the structural constraints dictating decision making at a lower level.

Codification, however, is predicated on the assumption that there is no change in the circumstances addressed at the higher, codified levels across decision making applications involving the use of the codebook or manual. Chaos may result in cases where this assumption is not met, as the principles involved in decision making at the higher levels have been explicitly excluded from consideration through the codification process. This serves to 'naturalize' (Barthes, 1973) the results of the use of higher level principles as codes of action. These codes appear, at the lower levels to be simply the 'natural' way to approach the decision problem and their history (the principles used to generate them) is lost from view.

Level 5

By analogy with relations between lower levels in our scheme, the content manipulated at level 5 comprises complete level 4 systems. What is the **form** though of the context within which they are located and may be manipulated by a decision maker who understands the principles underlying a level 5 problem structuring calculus? A clue can be gained from the description given in table 1 of the demand characteristics of tasks located at this level in a bureaucratic organization: problems are not dealt within a context wholly set from above and one can modify the boundaries of one's business within corporate policy. Here we have a link with Savage's (1954) notion of the **small world** within whose boundaries problem structuring activities and current knowledge representations are located. Toda(1976) has discussed the nature of these small worlds, and how their boundaries are closed according to the goals of the decision maker. This 'boundary setting' for a decision problem changes with changes in motivation, as does the nature of the structures the decision maker will consider requisite in handling any decision problem (c.f. Jungermann, von Ullardt and Hausman, 1983 and Humphreys et al, 1983).

Development of a structuring calculus at this level would mean the articulation of principles guiding the conditional closing of an open system (the unbounded "grand world" of possible views of the future) into the small world within which an individual's problem structuring activities take place. To our knowledge no example of such a calculus has yet been formalized.

Higher levels of abstraction

Above level 5 we move outside the scope of **individual** decision making as at level 6 the "small worlds" which define form at level 5 are themselves placed as content within structures generated at higher levels. We can understand the nature of knowledge representations at levels 6 to 10 by analogy with levels 1 to 5. At level 6, principles like seriation, transitivity and conservation are important in controlling operations, as at level 1, but the material on which they operate are complete small worlds, rather than unitary "facts". Also, level 6 decision making requires support at levels 1 to 5 in developing requisite conceptualizations of these small worlds. A

similar comparison may be made between level 7 and level 2, and so on.

The reason why we state that decision making at these higher levels of abstraction must of necessity be social rather than personal is that each small world manipulated as content at these levels marks for the individual located within it the boundary of the ways in which he or she can conceptualize scenarios of future personal experiences consequent on actions he or she might take. Decision making above level 5 is concerned with choosing between courses of action which, when supported at all lower levels, will lead to the generation and re-ordering of the cultural structures within which individuals find their identity. Few decision makers occupy roles with the scope and levels of organizational support which permit decision making tasks to be handled at these levels of abstraction. However, Jaques(1982) describes organizational roles and demand characteristics at levels 6 and 7. He claims to have identified a very few decision makers with the capability and organizational support allowing them to operate at level 8, but none above this.

Implications for the investigation of individual decision making

Given our analysis of levels of abstraction in the representation of decision problems, we conclude by examining its implications for the study of individual decision making under uncertainty through experiments conducted on intuitive inference and decision making within the "conversational paradigm" (see Kahneman and Tversky, 1982).

In analysing the results of such experiments, figure 1 may be applied in studying differences between subjects' responses and the output of a normative model at **any particular level** of abstraction. However, for the comparison procedure to have any validity, two conditions must be met: (i) comparisons must be made at one, and only one, pre-defined level of abstraction at a time, and (ii) for a comparison to be made at any particular level, there must be a common understanding about how the structure of the problem representation is to be fixed **at all higher levels** (that is, if the procedure in figure 1 were to be applied at these higher levels, there would be no difference in the two outputs).

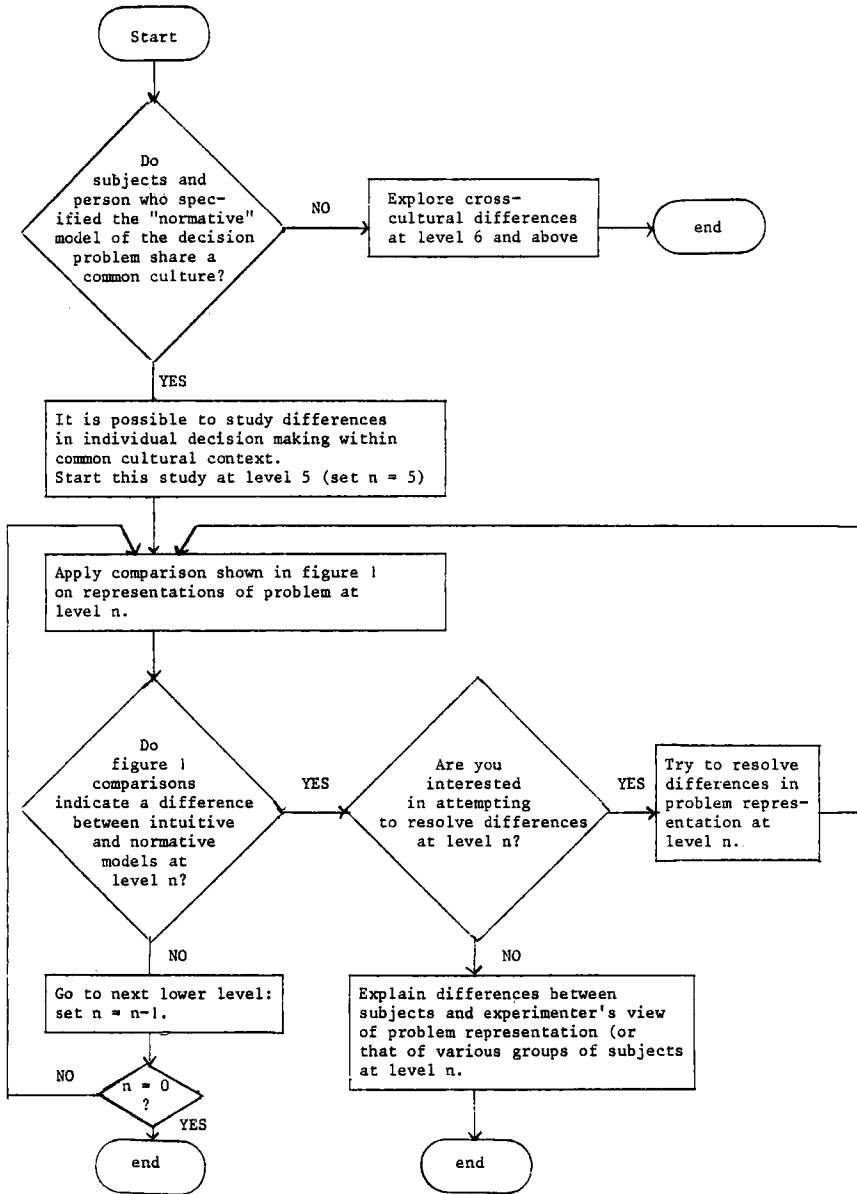
These two conditions together imply that we should start at the highest level at which there is a possibility of the existence of differences, and then work down, level by level, interpreting the differences in output comparisons shown in figure 1 at the highest level at which they occur.

The procedure involved is given in figure 5. As this paper has been concerned with studying differences in individual decision making, the highest level which is considered explicitly is level 5. Starting at level 5 assumes, however, that there is a common understanding at levels 6 and above. These levels are primarily concerned with structuring the cultural context within which individuals locate themselves, and, indeed, define their identities, topics outside the scope of this paper. Therefore, we start here by assuming a common understanding about the shared cultural context which is likely to exist if all concerned (experimenters, subjects, etc.) come from similar cultural backgrounds.

At level 5 we are concerned with possible differences in the way in which people set the bounds on the small worlds within which they may conceptualize the alternative futures that may follow from deciding upon any of the options under consideration. Humphreys(1982) has described how the way in which individuals set these bounds defines the limits of the 'background of safety' (Sandler and Sandler, 1978) within which scenarios of such possible futures are constructed and explored. Within these bounds, the way scenarios are articulated will depend upon the goals of the individual in defining characteristics of the (small) world he or she would, or would not, like to live in.

An example of an area where differences at this level come to the fore is public debates concerning conflicts over nuclear power. One side in such a debate (e.g. at a public inquiry) may define the 'acceptability' of nuclear power in terms of its safety where 'safety' is defined in terms of expected number of fatalities directly attributable to the reactor process in operation, whereas the other side may be principally concerned with (un)acceptability in terms of exploring scenarios concerning increased centralization of state power and security, and problems of future hazards from radio active waste (c.f. Gyorgi, 1981; Watson, 1981; von Winterfeldt and

Figure 5: Multi-level comparison procedure for investigating differences between "intuitive" and "normative" problem representations



Rios, 1980). In this case, seeking evidence which assumes that the difference lies at level 1 (e.g. "it is important to ascertain how safe a plant will be in terms of the best estimate of expected number of fatalities per year") will never resolve the differences between the parties. It cannot ever address them, as there is no common agreement about the scenarios of the future which need to be explored in ascertaining the 'acceptability' of what is thus conceptualized and which will define the themes to be structured at level 4 and below.

If however, consensus can be achieved at level 5, a common understanding concerning the bounds of the problem to be structured at that level may be defined and the comparisons shown in figure 1 can then be made at level 4, where we are interested in how individuals generate the structures which describe their view of the problem. If agreement is not reached on the **structural types** to be used in representing the problem, and how they are to be linked together (c.f. figure 4), then it is necessary to analyse the different linkages at this level. For example, consider the following questions, posed by Phillips and Boxall (1983) to subjects in an experiment within the 'conversational paradigm' (Kahneman and Tversky, 1982).

"Peter is considering installing a burglar alarm to protect his home. Which is more likely: (i) that Peter's home will be broken into before the end of next year if he installs a burglar alarm before the end of next year, or (ii) that Peter's home will be broken into before the end of next year if he does not install a burglar alarm before the end of next year?"

Phillips and Boxall (1983) report an analysis of the problem representations actually used by subjects (obtained through a 'think aloud' technique, following Feldman, 1963 and Svenson, 1974) in deciding how to answer this and eleven other questions developed from those originally posed to subjects by Tversky and Kahneman. They found that 58% of the subjects elaborated the question in a different way from the linkage of one binary event followed by another assumed by Tversky and Kahneman (1980) in discussing the external 'diagnostic' perspective subjects may use in arriving at an answer. In some cases, subjects role-played Peter or the burglar, taking an internal

perspective when the problem is represented through an act-event sequence, instead of an event-event sequence.

Accounting for differences between subjects who adopted an 'external' perspective, and those who adopted an 'internal' perspective thus required an explanation centred on alternative level 4 problem representations. However, not all the questions studied by Phillips and Boxall led to differences which had to be accounted for at level 4. For some, the information provided to subjects was found to be sufficient to ensure a 'common understanding' of the problem representation at this level.

Achievement of a common understanding at level 4 allows examination of differences in the way in which individuals elaborate their representations of a decision problem within a particular structural type represented at level 3. 'Social decision making' schemes located at this level have been described by Edwards (1977) and von Winterfeldt and Rios (1980), among others. Each of these reports describes cases where consensus between various individuals was assumed or engineered at levels 4 and 5, and differences displayed within a problem representation of a single structural type: a multiattribute utility decomposition comprising all attributes which **any** of the parties considered relevant. Differences between parties can then be expressed by different vectors of weights applied within the overall decomposition (assigning a zero weight to any part of the decomposition effectively deletes it from consideration as would be the case where the attributes in the 'social decision making' decomposition were not present in a particular individual's multiattribute utility structure, characterizing his or her level 3 view of the problem).

Berkeley and Humphreys (1982) and Phillips and Boxall (1983) describe some alternative elaborations which may be used by subjects presented with decision problems by experimenters using the 'conversational paradigm'. In the scheme we describe here, these alternative elaborations represent differences in ways of conceptualizing the decision problems at level 3. Here, the appropriate explanation has to be in terms of accounting for the differences in elements in the structure (e.g. intervening acts, events or attributes of consequences) and in how they are linked together.

When, finally, we can ensure that there is a common understanding between subjects and experimenter concerning the structure of the problem representation at level 3, it then becomes possible to make the comparisons shown in figure 1 at level 2. Now we can look, for example, for evidence whether subjects are 'normatively consistent' or biased towards causal reasoning (c.f. Tversky and Kahneman, 1980) in the way in which they interpret the diagnostic significance of evidence presented in sequences like that given in the 'burglar alarm' question we discussed above.

In summary, we share with Amarel (1968) the view that

"The act of choosing a representation for a problem involves the specification of a space where the search for solution can take place. Such a specification involves the choice of a language - and its use - for expressing problem conditions, properties of solutions, and knowledge of regularities in the search space. In addition, it involves the formulation of a schema according to which problem conditions and knowledge about properties of the space can be used effectively in the solution seeking procedures. Such a schema can be as simple as a set of local moves or as elaborate as a global plan.." (Amarel, 1968; p.94).

In this paper we have attempted to explore this space, defining general conditions for the formulation of such a schema and illustrating how at least five qualitatively different levels of abstraction must be encompassed in this enterprise.

References

- Amarel, S.: 1968. Problems of representation in Artificial Intelligence. In: von Foerster, H., White, J.D., Peterson, L.J. & Russell (eds), *Purposive Systems*. Washington. Spartan Books.
- Anderson, J.R.: 1976. *Language, memory and thought*. Hillsdale. Erlbaum.
- Bar-Hillel, M.: 1982. Studies of representativeness. In: Kahneman, D., Slovic, P. & Tversky, A. (eds), *op.cit.*
- Barthes, R.: 1973. *Mythologies*. St Albans. Paladin.

- Berkeley, D.: 1982. The meaning of rationality when conceptualizing decision problems. Occasional paper. London. Department of Social Psychology, London School of Economics and Political Science.
- Berkeley, D. & Humphreys, P.C.: 1982. Structuring decision problems and the 'bias heuristic'. *Acta Psychologica*, 50, 201-252.
- Bernouilli, D.: 1738. *Specimen theoriae novae de mensura sortis*. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, 175-192 (trans. by L.Sommer, 1954. *Econometrica*, 22, 23-36).
- Bonczek, R.H., Holsapple, C.W. & Winston, A.B.: 1981. *Foundations of decision support systems*. New York. Academic Press.
- Bronner, F. & de Hoog, R.: 1983. Non-expert use of a computerized decision aid. In: Humphreys, P.C., Svenson, O. & Vari A.(eds), *op.cit.*
- Brown, R.V.: 1978. Heresy in decision analysis: Modeling subsequent events without rollback. *Decision Sciences*, 9, 543-554.
- Brown, R.V., Kahr, A.S. & Peterson, C.R.: 1974. *Decision analysis for the manager*. New York. Holt, Rinehart & Winston.
- Brown, R.V. & Lindley, D.V.: 1982. Improving judgement by reconciling incoherence. *Theory and Decision*, 14, 113-132.
- Brown, R.V. & Ulvila, J.W.: 1977. *Selecting analytic approaches for decision situations (Revised edition)*. Vol.1 : Overview of the methodology. Technical report TR77-7-25. Maclean, Va. Decisions and Designs Inc.
- Bunge, M. : 1969. *Metaphysics, Epistemology and methodology of levels*. In: Whyte, L.L., Wilson, A.G. & Wilson, D.(eds): *Hierarchical Structures*. New York. American Elsevier.
- Chomsky, N.: 1957. *Syntactic structures*. The Hague. Mouton.
- Edwards, W.: 1954. The theory of decision making. *Psychological Bulletin*, 51, 380-417.
- Edwards W.: 1977. Use of Multiattribute utility analysis for social decision making. In: Bell, D.E., Keeney, R.C. and Raiffa, H. (eds), *Conflicting objectives in decisions*. New York. Wiley.
- Edwards, W., Lindman, H. & Phillips, L.D.: 1965. Emerging technologies for making decisions. In : Newcomb, T.M.(ed), *New directions in Psychology*, II. New York. Holt, Rinehart and Winston.
- Elster, J.: 1979. *Ulysses and the Sirens: studies in rationality and*

irrationality. London. Cambridge University Press.

Feldman, J.: 1963. Simulation of behaviour in the binary choice experiment. In: Feldman, J. & Fergenbaum (eds), Computers and thought. New York. McGraw Hill.

Fillmore, C.J.: 1968. The case for case. In: Bach, E. & Harms, R.T. (eds): Universals in linguistic theory. New York. Holt, Rinehart & Winston. (pp.1-88)

Fischhoff, B.: 1982. Debiasing. In: Kahneman, D., Slovic, P. & Tversky, A.(eds), op.cit.

Gyorgy, A. & friends: 1979. No Nukes: Everyone's guide to Nuclear Power. Boston, Mass. South End Press.

Hogarth, R.M., Michaud, C. & Mery, J.L.: 1980. Decision behavior in urban development: A methodological approach and substantive considerations. Acta Psychologica, 45, 95-117.

Howell, W.C. & Burnett, S.A.: 1978. Uncertainty measurement : a cognitive taxonomy. Organizational Behavior and Human Performance, 22, 45-68.

Hovland, C.I.: 1952. A 'communication analysis' of concept learning. Psychological review, 59, 461-472.

Humphreys, P.C.: 1977. Applications of Multiattribute Utility Theory. In: Jungermann, H. & de Zeeuw G.(eds), Decision making and change in human affairs. Dordrecht. Reidel.

Humphreys, P.C.: 1982. Value structures underlying risk assessments. In: Kunreuther, H. (ed.). Risk: a seminar series. Laxemburg, Austria. International Institute of Applied Systems Analysis.

Humphreys, P.C., Larichev, O.I., Vari, A. & Vecsesnyi, J.: 1983. Comparative analysis of use of decision support systems in R&D decisions. In: Sol, H. G. (ed): Processes and tools for decision support. Amsterdam. North Holland.

Humphreys, P.C. & McFadden, W.: 1980. Experiences with MAUD: aiding decision structuring versus bootstrapping the decision-maker. Acta Psychologica, 45, 51-69.

Humphreys, P.C., Svenson, O. & Vari, A. (eds): 1983. Analysing and aiding decision processes. Amsterdam. North Holland.

Humphreys, P.C. & Wisudha, A.: 1982. MAUD4. Decision Analysis Unit Technical report 82-5. London. London School of Economics and Political Science.

- Humphreys, P.C., Wooler, S. & Phillips, L.D.: 1980. Structuring decisions: the role of structuring heuristics. Decision Analysis Unit, Technical report 80-1. Uxbridge, England. Brunel University.
- Jaques, E.: 1976. A general theory of Bureaucracy. London. Heinemann.
- Jaques, E.: 1982. Free enterprise, fair employment. London. Heinemann.
- Jaques, E., Gibson, R.O. & Isaac, D.J.: 1978. Levels of abstraction in logic and human action. London. Heinemann.
- Johnson, D.M.: 1972. A systematic introduction to the psychology of thinking. New York. Harper & Row.
- Jungermann, H., von Ullardt, I. & Hausman, L.: 1983. The role of goal in generating actions. In: Humphreys, P.C., Svenson, O. & Vari, A.(eds), op. cit.
- Kahneman, D., Slovic, P. & Tversky, A. (eds) : 1982. Judgement under uncertainty: Heuristics and Biases. New York: Cambridge University Press.
- Kahneman, D. & Tversky, A.: 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 47, 263-291.
- Kahneman, D. & Tversky, A.: 1982. On the study of statistical intuitions. *Cognition*, 11, 123-141.
- Kneppreth, N.P., Gustafson, D.H., Leifer, R.P. & Johnson, E.M.: 1974. Techniques for the assessment of worth. Technical paper 254 (AD784629). Arlington, Va.. Army Research Institute for the Behavioral and Social Sciences.
- Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A.: 1971. Foundations of measurement (Vol 1). New York. Academic Press.
- Krantz, D.H.: 1972. Measurement structures and psychological laws. *Science*, 175, 1427-1435.
- Leach, E.R.: 1964. Anthropological aspects of language. In: Lenneberg, E.H.(ed). New directions in the study of language. Boston, Mass. Institute of Technology Press, 23-63.
- Lichtenstein, S.C., Fischhoff, B. & Phillips, L.D.: 1982. Calibration of probabilities: the state of the art to 1980. In: Kahneman, D., Slovic, P. & Tversky, A.(eds), op.cit.
- McCosh, A.M. & Scott Morton, M.S.: 1978. Management decision support systems. London. MacMillan.

- Newell, A. & Simon, H.A.: 1972. Human problem solving. Englewood, N.J. Prentice Hall.
- Nisbett, R.E. & Ross, L.: 1980. Human inference: Strategies and shortcomings of social judgement. Englewood Cliffs, N.J. Prentice Hall.
- Parkes, E.M.: 1971. Psychosocial transitions: a field for study. *Social Science and Medicine*, 5, 101-119.
- Phillips, L.D.: 1982. Requisite decision modeling: a case study. *Journal of Operations Research Society*, 33, 303-311.
- Phillips, L.D.: 1983. A theoretical perspective on heuristics and biases in probabilistic thinking. In: Humphreys, P.C., Svenson, O. & Vari, A. (eds), *op.cit.*
- Phillips, L.D. & Boxall, J.: 1983. A problem structuring analysis of heuristics and biases. Decision Analysis Unit. Technical Report 83-1. London. London School of Economics and Political Science.
- Piaget, J.: 1971. *Structuralism*. London. Routledge and Kegan Paul.
- Piaget, J.: 1972. *The principles of genetic epistemology*. London. Routledge and Kegan Paul.
- Piaget, J.: 1977. *The development of thought: equilibration of cognitive structures*. Oxford. Basil Blackwell.
- Sandler, J. & Sandler, A.M.: 1978. On the development of object relations and affects. *International Journal of Psychoanalysis*, 59, 285-296.
- Savage, L.J.: 1954. *The foundations of statistics*. New York. Wiley.
- Svenson, O.: 1974. A note on think aloud protocols obtained during the choice of a home. Reports from the Psychological laboratories. University of Stockholm, No 421.
- Swets, J.A., Tanner, W.P., Jr & Birdsall, T.G. : 1961. Decision processes in perception. *Psychological Review*, 68, 301-340.
- Thorngate, W.: 1980. Efficient decision heuristics. *Behavioral Science*, 25, 219-225.
- Toda, M.: 1976. The Decision Process: a perspective. *International Journal of General Systems*, 3, 79-88.
- Tversky, A. & Kahneman, D.: 1974. Judgement under uncertainty : heuristics

and biases. *Science*, 185, 1124-1131.

Tversky., A. & Kahneman, D.: 1980. Causal schemas in judgements under uncertainty. In: Fishbein, M. (ed.). *Progress in Social Psychology*. Hillsdale, N.Jersey. Erlbaum.

Tversky, A. & Kahneman, D.: 1981. The framing of decisions and the psychology of choice. *Science*, 211, 453-458.

Vari, A. & Vecsenyi, J.: 1983. Decision analysis of industrial R&D problems: Pitfalls and lessons. In: Humphreys, P.C., Svenson, O. & Vari, A.(eds), *op. cit.*

Watson, S.R.: 1981. On risks and acceptability. *Journal of the Radiological Society*, 1, 21-25.

von Winterfeld, D.: 1980. Structuring decision problems for decision analysis. *Acta Psychologica*, 45, 71-94.

von Winterfeldt, D. & Rios, M.: 1980. Conflicts about Nuclear safety: a decision theoretic approach. Social Science Research Institute. Los Angeles. University of Southern California.

Zimmer, A.: 1983. Verbal versus numerical processing of subjective probabilities. In: this volume.

This Page Intentionally Left Blank

VERBAL VS. NUMERICAL PROCESSING OF SUBJECTIVE PROBABILITIES

ALF C. ZIMMER

*Universität Münster
Psychologisches Institut
F.R. Germany*

The problem of deciding between different alternatives with equally positive outcomes has an apparently trivial solution: take the one where the positive outcome is most probable. That this solution is not as trivial as it seems to be can be seen in the fact that it was not until 1660 for the concept of probability to be treated mathematically (HACKING, 1975), thereby providing solutions for how to decide in the case of uncertain events.

As early as 1713 JACQUES BERNOULLI has made the distinction between probabilities - later defined by LAPLACE as the relative frequencies of successes - and subjective "degrees of confidence". By means of this distinction it became possible to account for inconsistencies or irrationalities in human decisions. The course of action taken was assumed to depend on these subjective "degrees of confidence" instead of the objective probabilities as derived from long-run experience. Practical as well as philosophical considerations then led to various attempts to define subjective probability formally (e.g. DE FINETTI, 1937) and to prescribe procedures of determining whether numerical estimates of probabilities given by human raters fulfil the axioms of subjective probabilities (The system of axioms most often investigated is the one due to SAVAGE, 1954). A thorough critique of this approach can be found in SLOVIC & TVERSKY (1974) and SUPPES (1974). LINDLEY (1974) adds a further discussion from the point of view of Bayesian statistics. The core of both SUPPES' and TVERSKY's critique is that human decision makers and probability raters markedly and systematically deviate from what is prescribed by SAVAGE's theory, especially concerning the axiom of indifference.

From the point of view of human information processing the framework for investigating this problem is given in table 1.

Table 1

Human information processing depends on:

1. Modes of representation

verbal

(i) propositional numerical

generally symbolic

imaginal

(ii) analogue

automatic frequency monitoring

these modes differ in: - the transformations which can be applied.

The constraints on these representations can be "strict" or "elastic".

- the mental work they impose on the information processing capacity

2. Available procedures

(i) rules (e.g. grammars, arithmetics, "Gestalt laws" in perception, etc.) these interact with the above mentioned applicable transformations and are therefore context specific.

Less context specific are

(ii) heuristics, which can be regarded as tools for narrowing down the number of possible candidates among the transformations in the case of insufficient information

Least context specific and therefore applicable in situations of information overload are

(iii) rules of thumb, which can be applied almost without any constraints. They provide "quick and dirty" procedures for the reduction of information so that the processing overload is relieved. (e.g. "Take whatever comes to your mind first".)

Most investigations of subjective probabilities have been much more concerned with the underlying procedures (see bottom half of table 1) than with the modes of representation (see upper half of table 1). These investigators have implicitly assumed that information is stored propositionally and that after retrieval it can be stated numerically. Any inconsistency between information intake (the objective side) and the numerically expressed subjective probability (the subjective side) is then ascribed to the procedures. That is, the subject is presumed to have chosen an inappropriate algorithm. Examples for such proposed procedural mistakes are overconfidence (see Figure 1), conservatism (i.e. sticking to an initial appraisal of a situation in spite of new information available for revision), and negligence of the regression effect (i.e. the implicit assumption of a perfect correlation between the predictor variable and the criterion).

The approach taken here begins by asking how uncertain events might be represented internally. That is, are they represented in (i) a verbal propositional mode, (ii) a numerical propositional mode, or (iii) in an analogue mode of automatic frequency monitoring (see the next chapter). Thus, what is questioned here is not the adequacy of human judgments concerning uncertainty, but rather the individuals' ability to express numerically what is internally represented.

Coping with uncertainty is ubiquitous in humankind (WRIGHT and PHILLIPS, 1980) and expressions for different degrees of certainty can be found in most languages. Yet, as reported above it took until the 17th century to develop these concepts mathematically. Even then it took place only in the European culture where it was motivated by practical problems; such as gambling and insurance. It seems unlikely that the mathematically appropriate procedures with numerical estimates of uncertainty have become automatized since then. It is more likely that people handle uncertainty by customary verbal expressions and the implicit and explicit rules of conversation connected with them. In the research reported here I therefore start from the analysis of meaning of common verbal expressions for uncertain events. These expressions are interpreted as possibility functions (ZADEH, 1978) and the procedures applicable to them (rules as well as heuristics and biases) are modelled in the framework of possibility theory. Since this theory allows

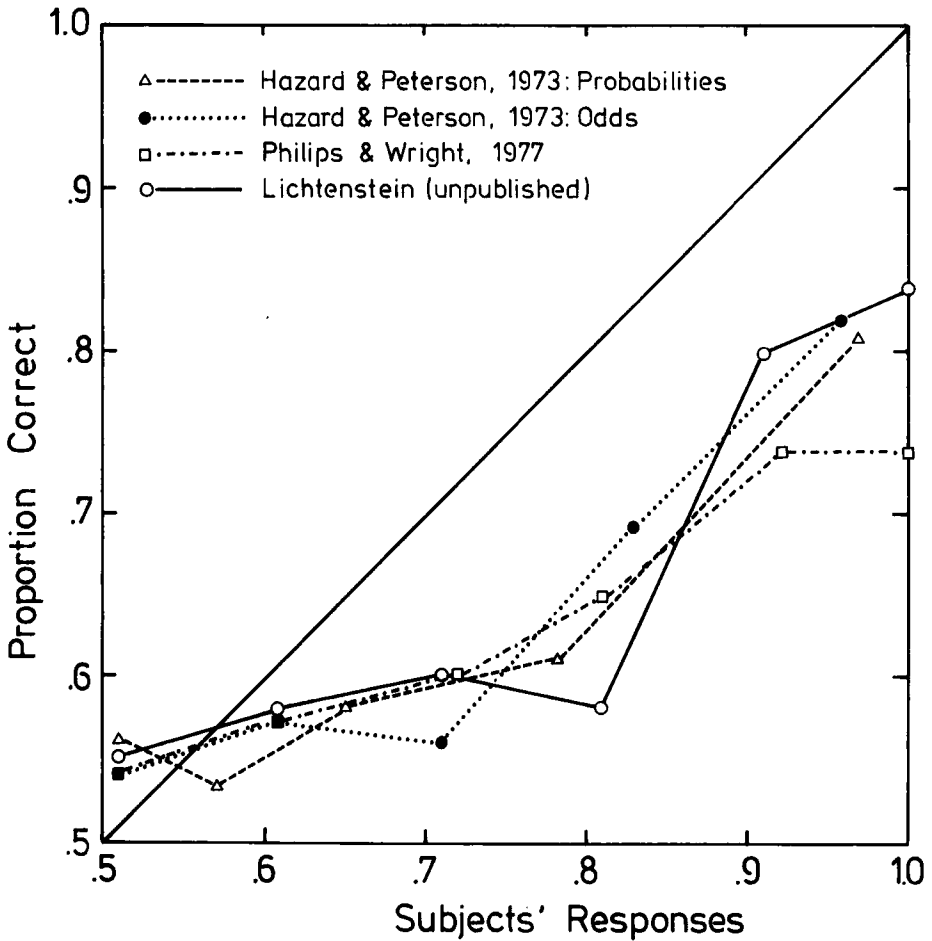


Figure 1. Calibration curves adapted from LICHTENSTEIN, FISCHHOFF, and PHILLIPS (1982)
Overconfidence is shown by the fact that in the upper part of the graph the subjective probability estimates are significantly higher than the objective relative frequencies (probabilities).

for a numerical interpretation by means of determining the elastic constraints on the usage of such expressions, the results gained by interpreting verbal expressions of uncertainty as possibility functions can be compared to the results of the above mentioned studies, where subjects provide the numerical expressions themselves.

The Automatic Monitoring of Frequencies

Before developing this theory in detail for the purposes of our study it seems to be worthwhile to ask how people gather the knowledge which is reflected in the verbal expressions for the probabilities of uncertain events. One can assume that for repetitive events e.g. the outcomes of ballgames or the daily weather, they monitor the frequency of outcomes automatically and that they revise their knowledge accordingly. HINTZMAN, NOZAWA, and IRMSCHER (1982) have studied this automatic monitoring of frequencies and they report that it is not interfered with by numerical processing. From their result they conclude that this information is stored in a non-numerical analogue mode (see item 1 (ii) in Table 1). Such automatic monitoring of frequencies seems to be a plausible candidate for the mode of representation underlying the generation of judgments concerning subjective probabilities.

In an experiment I have tried to determine whether the frequencies of more than one unattended stimulus attribute can be monitored in order to determine the degree of automaticity and mutual interference of multiple frequency judgments. In the experiment 150 color slides depicting either landscapes or buildings were presented to 80 Introductory Psychology students in a probability learning task, in which they had to predict the content - building vs. landscape - of each upcoming slide. Afterwards they were asked for the relative frequencies of two presumably attended features of the slides, orientation and form, which were uncorrelated with the content of the slides. The question format was either a numerical scale, on which subjects had to mark their frequency estimate, or a sequence of verbal expressions for relative frequencies, out of which they had to pick the best fitting one. For the exact experimental design and the results (mean squared errors) see Table 2.

The results show (i) the unattended features differed in saliency (orientation was easier to monitor than form), (ii) the verbal judgments were slightly better than the numerical ones, (iii) in all cases the second frequency judgments were worse than the first ones, but more important, the impairment was less in all cases where one of the judgments was verbal and it was least when both were verbal. From these results it seems plausible to conclude, first, that more than one unattended variable can be automatically monitored even if they are closely related, but the mode of probing this

Table 2

The verbal judgments have been transformed into numerical values by assuming that the labels partition the range (0% - 100%) equally.

mean squared errors	mode of judgment				
	<u>verbal</u>		<u>numerical</u>		
	other position verbal judgm.	numerical j.	other position verbal judgm.	numerical j.	
Form	position 1	10.1	9.8	12.3	13.1
	position 2	15.5	16.2	17.4	22.0
Orientation	position 1	5.7	7.1	8.1	7.9
	position 2	9.2	8.7	13.8	19.5

knowledge influences the precision of judgments. Second, if more than one judgment has to be made there is inference between them, but the amount of inference depends on the modes of judgments. This apparent superiority of the verbal mode leads to the tentative interpretation that the verbal mode for representing knowledge is able to process information more effectively than the numerical mode, because the rules applicable to processing verbal propositions have been acquired earlier and are therefore more automatized. This would explain why they require less mental processing capacity and consequently are less prone to interference from simultaneous processes. If the over-

load of the mental processing capacity necessitates the application of heuristics, and if in turn biases in human judgment can be traced back to mistaken applications of heuristics (for an overview see NISBETT and ROSS, 1980), the following conjecture seems to be plausible: Any mode of judgment imposing less mental work load should be more valid than one requiring more mental processing capacity. This conjecture motivated the following experiments, in which I have looked for possible ways of making the information processing easier in tasks which imply the analysis of uncertain events. In order to analyze the efficiency of different modes of representation in information processing it is necessary to devise an interpretation for the meaning of verbal expressions which can be translated into numerical values. Fuzzy set theory provides a tool for such interpretation by means of numerically stated elastic constraints for the applicability of verbal expressions.

A Fuzzy Set Theoretical Approach to Subjective Probabilities

According to fuzzy set theory the meaning of concept " x " in a universe of discourse " U " can be modelled by the possibility function for " x " in " U ", which indicates for which states in " U " the concept " x " fits, for which it can be possibly applied, and for which it does not fit at all. Figure 2 depicts the possibility functions for two concepts as well as the basic two binary operations:

$$(i) \text{ the intersection } \quad \underline{A} \cap \underline{B} = \underset{U}{\text{MIN}} (f_{\underline{A}} ; f_{\underline{B}})$$

$$(ii) \text{ the disjunction } \quad \underline{A} \cup \underline{B} = \underset{U}{\text{MAX}} (f_{\underline{A}} ; f_{\underline{B}})$$

The version of fuzzy set theory applied here, especially concerning the form of possibility functions, deviates from approaches where the exact form of these functions is determined either theoretically (for an overview see KAUFMANN, 1975, or DUBOIS and PRADE, 1980) or empirically (e.g. HERSH and CARA-MAZZA, 1976; ZIMMER, 1980, 1982 a). The appropriateness of these specific membership functions as parts of the mental representation of vague concepts

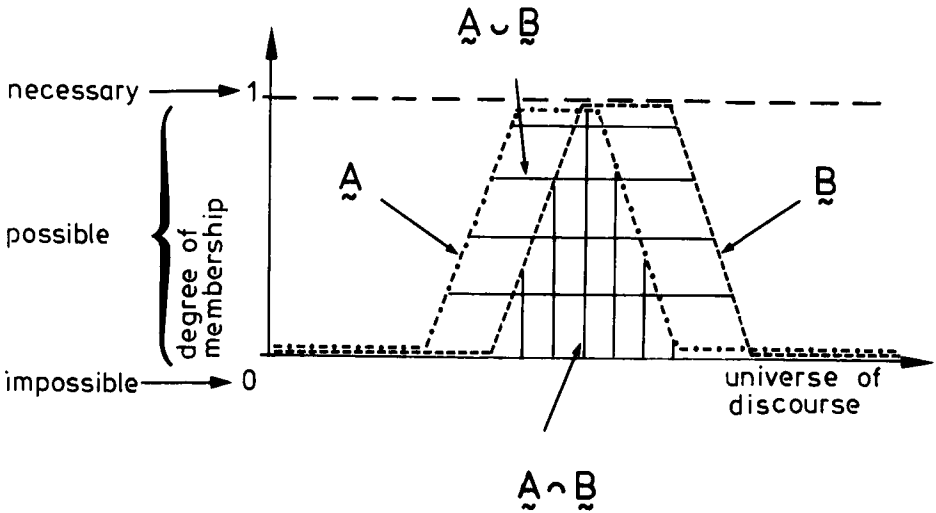


Figure 2. Possibility distributions and binary operations upon them

is at least doubtful. FREKSA (1981, 1982) has shown that for practical purposes such specification is unnecessary. By restricting the possibility to the values: "necessary", "possible", and "impossible", one can avoid the debated question of how exact the information about one's own knowledge can be (cf. NISBETT and WILSON 1977) without losing the amount of specificity necessary to apply fuzzy set theory to verbal concepts. The meanings of these concepts is then given by the elastic constraints imposed upon them by the possibility functions.

Modelling the meanings of verbal labels for relative frequencies is straightforward in this framework. The universe of discourse is the unit interval and the regions of applicability, possible applicability, and inapplicability can then be determined empirically. The resulting possibility functions for a given set of verbal expressions is depicted in Figure 3. The spacing of the possibility functions in the unit interval is reminiscent of WITTE's (1960) treatment of verbal judgments stabilized in memory over time. I (ZIMMER, 1980) have shown that such equidistant and equally shaped categories are optimal from the point of view of language pragmatics, but it has to be kept

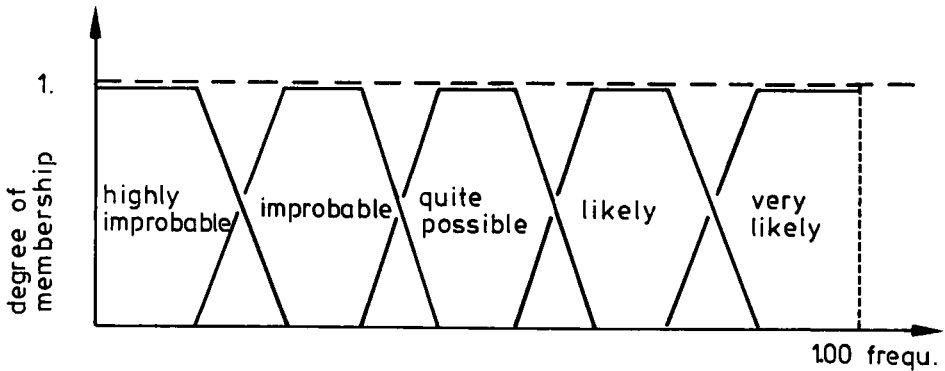


Figure 3. Possibility distributions for the meaning of verbal expressions for uncertainty for a subject using 5 labels

in mind that this communicability constraint is a simplifying assumption, which is not a prerequisite for this kind of modelling. WITTE's approach, however, implies two questionable assumptions: that the verbal labels are uniquely ordered for all subjects and that all subjects are equally familiar with the labels provided by the experimenter.

Empirical Determination of the Fuzzy Meanings of Verbal Expressions

In the preliminary step for my experiment I observed the subject's usage of expressions for uncertain events and asked them afterwards to order the used labels according to frequency. It turned out that the subjects differed markedly in verbal labels used and moreover in the serial order of the same labels used by different subjects. For instance some subjects reported the increasing serial order "possible, likely, probable" whereas others interchanged "possible" and "likely" or "likely" and "probable" (WALLSTEN, note 1, reports similar results). In the following it will be assumed that subjects used these expressions in the way they reported. As it will turn out later this assumption is in line with the empirical results.

The elicitation of the verbal labels that subjects were most familiar with and used most frequently, revealed that most subjects used merely five or six different expressions (see Table 3). A total of twelve different labels was

Table 3

number of labels for uncertain events	4	5	6	7	8
number of subjects	2	87	55	5	1

found. The empirical definitions of the meanings had to be done individually because of the interindividual differences. From the point of view of language pragmatics these individual differences in using the same expressions are a puzzling aspect because they contradict the assumed communicability constraint and cast into doubt the pragmatic value of such verbal judgments. From the interviews with the subjects in the debriefing sessions I learned that these marked individual differences were partially due to the compositions of our sample of subjects: 30 students in psychology, 30 students in business and economy, and 90 drafted soldiers of the German Bundeswehr. Whereas the first two groups had finished courses in statistics and probability theory, in the last group only a small proportion had studied probability theory in high-school. When asked what they would do if they realized that somebody else might misunderstand their verbal frequency expressions, most subjects reported that in such cases they would use the reference to generally known standard situations as a means to resolve this misunderstanding with an analogy. For instance they might point out that a certain event is as probable as winning in the German Lottery.

The empirical determination of the meanings for the individually used verbal expressions is done in the following steps:

- (i) survey of recurrently used labels for relative frequencies
- (ii) test of equivalence for labels (Do some of the words mean the same for you?)

- (iii) determination of the ranges for the labels, $s(k)$, assuming that the labels partition the overall range equally
- (iv) sequential estimation of the empirical meanings for the labels used by the subjects under examination:

The first step in the experimental determination of the meanings for the verbal expressions used by the subjects consisted in determining the starting point for the estimation procedure. This was done by taking the midpoint of the range in case of an uneven number of the verbal labels, that is, 50%, or in the case of an even number of categories a starting position was taken slightly moved to the lower or upper part of the range, that is, either 33% or 67%. In order to represent objective relative frequencies for the calibration of the individually used verbal expressions, I showed a window on a computer-driven tv-screen which contained up to 6144 dots (100%). The number of the white dots spaced randomly in the window in relation to the size of the screen represented the objective relative frequencies. After the starting configuration had been shown the subject had to give a verbal expression for the chance that any point inside of the window would be covered by a white dot, when the same frequency of dots were presented again. After the subjects had given their answers the next relative frequency of white dots was determined according to a modified ROBBINS-MONROE procedure with an ever decreasing step size as long as the subjects gave the same verbal label. When the subject changed the label the direction of change for the relative frequencies was changed, too. After more than 4 switches in direction, a new starting position was taken in another part of the range according to the number of verbal expressions used by the subject in question. This recursive procedure was repeated until the boundaries for each verbal label had been determined. The boundaries, that is, the areas where subjects switched from one label to another, were interpreted as the elastic constraints characterizing the meaning of the verbal expression.

A typical result of this experimental procedure is shown in Figure 3.

The results (e.g. Figure 3, label "likely") can be interpreted in the following way: for the subject under consideration the label "likely" fits for

events with the relative frequency of occurrence between .65 and .75. It might be applied too in the intervals .55 - .65 and .75 - .85 (possible applicability), and it is not a permissible expression for events outside of this interval.

The Modelling of the Availability Bias in this Model

I have argued elsewhere (ZIMMER, 1982 b) that the context specific meanings of quantifiers in colloquial English can be decomposed into knowledge about the occurrence of quantified statements in these contexts and into context free meanings of quantifiers. This separability of contextual and general meaning might explain why my subjects are apparently able to distinguish between at least four quantifiers, whereas BEGG (1982), who does not take into account the context, concludes from his data that normal language relies on only three quantifiers. This separability can also be used to model the availability bias in the model developed in this paper. It has been shown (for an overview see TVERSKY and KAHNEMAN, 1973) that judging the relative frequency of a class of events, such as the probability of dying due to an aircraft accident, subjects take into account not only the occurrence of this event but also the difficulty or ease they have in coming up with a typical example for this class of events; the more available information about the event there is, the more the probability of occurrence is overestimated. This availability bias can be accounted for in our model by the conjunction of the fuzzy meaning of each verbal expression under consideration and a scope function representing the influence of availability. In Figures 4, 5, and 6 scope functions for high, moderate, and low availability are shown. For these scope functions only monotonicity is assumed; the exact shape is to be determined empirically, e.g. by measuring the association speed or a similar variable. The conjunction of these scope functions with the possibility functions for the meanings of the verbal labels leads, after renormalization, to possibility functions of the biased judgments (see figures 7, 8, and 9). The comparison of the biased with the context free possibility judgments reveals that the model predicts an upward shift of the high uncertainty labels in the case of high availability and a downward shift of the low uncertainty labels in the case of low availability, whereas the labels on the other end of the range remain virtually unchanged. This effect is in line with experi-

mental results (e.g. those of SLOVIC, FISCHHOFF, and LICHTENSTEIN, 1979). If the postulated separability of the fuzzy meanings and the availability bias proves to be empirically valid, this model can have fruitful applications in methods for estimating subjective probabilities, because it provides a procedure for debiasing (for different approaches to debiasing see FISCHHOFF, 1982).

Scope functions

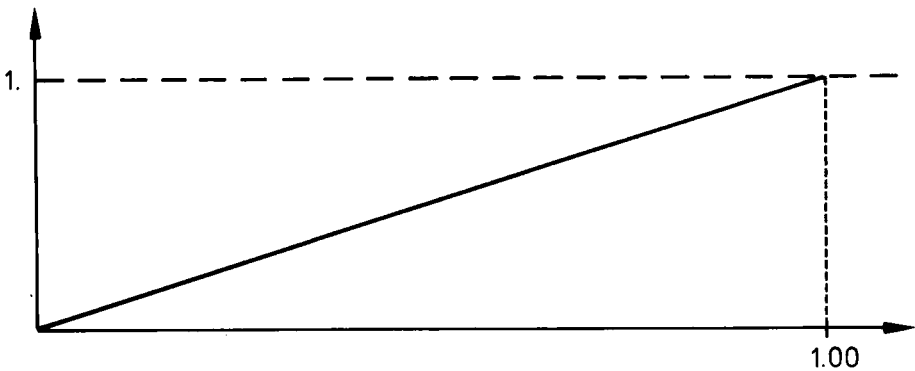


Figure 4. Scope function for a "high availability bias"

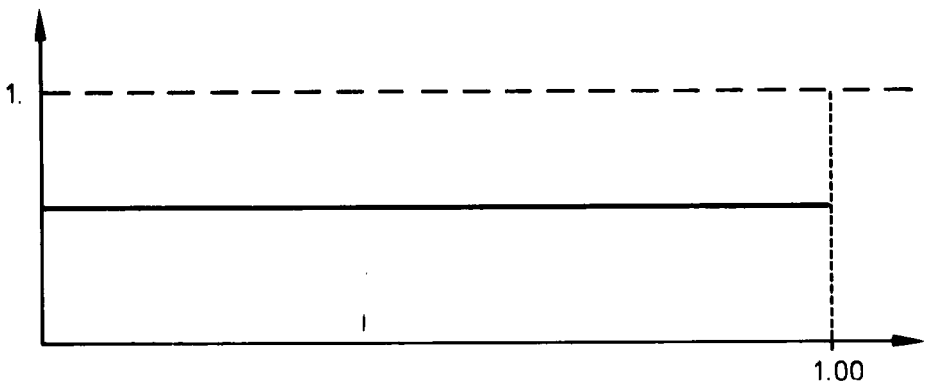


Figure 5. Scope function for a "moderate availability bias"

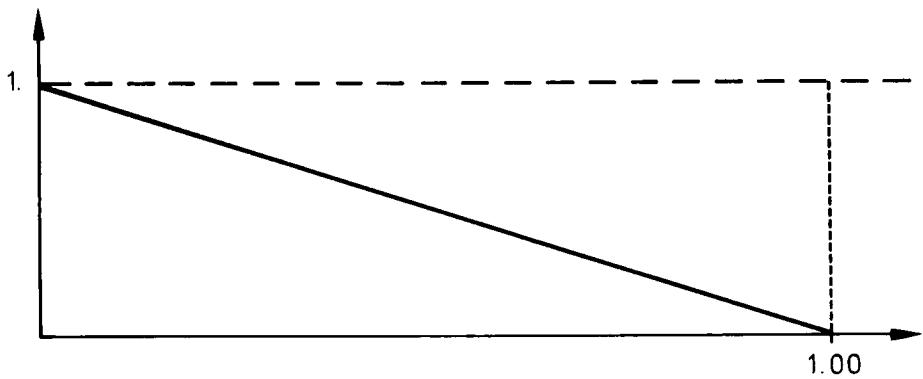


Figure 6. Scope function for a "low availability bias"

Influence of availability on categories

----- $x \cap \text{bias}$
—— normalized ' $x \cap \text{bias}$ '

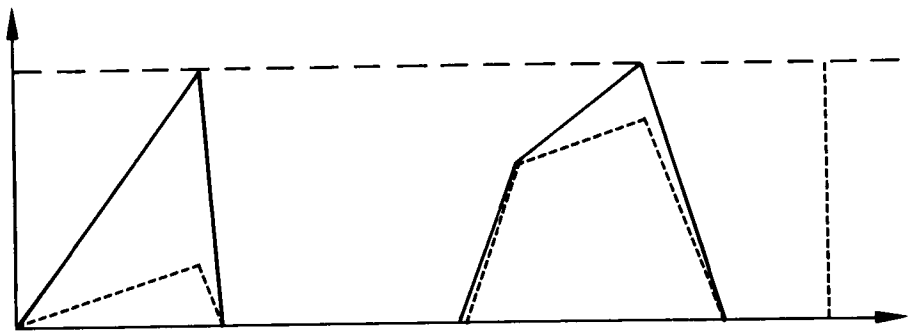


Figure 7. Influence of high availability on the meaning of "highly improbable" and "likely" in Figure 3

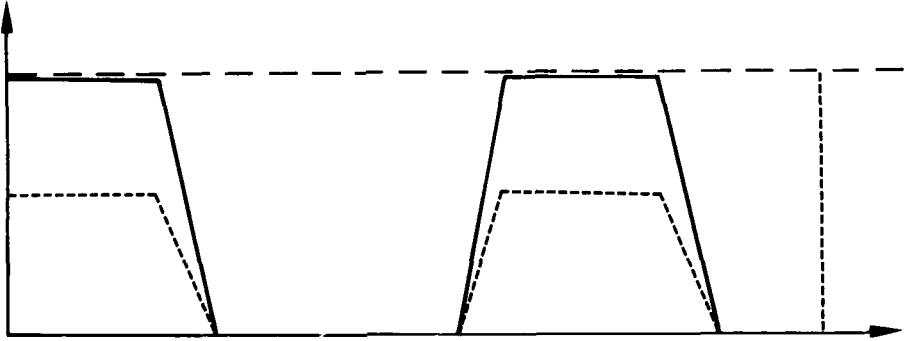


Figure 8. Influence of moderate availability on the meaning of "highly improbable" and "likely" in Figure 3

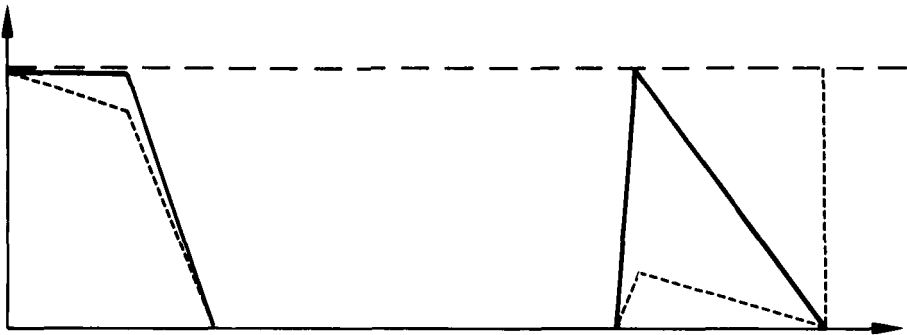


Figure 9. Influence of low availability on the meaning of "highly improbable" and "very likely" in Figure 3

Up to this point I have implicitly assumed that the possibility functions for the verbal labels are well calibrated, which is tautologically true for the relative frequencies of white dots on a tv screen. It can also be shown for tasks usually undertaken to test calibration, where subjects answer questions about almanac knowledge and estimate immediately afterwards the subjective probability of their answer being correct (for an overview see LICHTENSTEIN, FISCHHOFF, and PHILLIPS, 1982).

Calibration of Verbal Expressions for Probabilities

In order to separate the variables "difficulty of the item" and "of the subject", both of which influence the probabilities of correct answers, the conjoint measurement approach of RASCH (1966) is applied, which provides independent estimates of the ability of a subject, $\xi(i)$, and the difficulty of an item j , $\delta(j)$. These estimates in turn allow us to assess the probability that a subject i solves item j , according to the following formula:

$$P(+|\xi_i; \delta_j) = \frac{e^{\xi_i - \delta_j}}{1 + e^{\xi_i - \delta_j}}$$

MAY (note 2) has suggested this approach to calibration which takes into account individual differences. The specific difficulty of a given item j for a subject i is the difference between the difficulty of the item and the ability of the subject. On this scale the relative frequencies of the labels which the subjects use to characterize their probability of success, can be compared with the estimated objective probabilities.

A test consisting of 150 political knowledge items was administered to 90 drafted soldiers of the German Bundeswehr. The items were in an open response format with subjects first answering the item and immediately afterwards giving their verbal expressions for the probability of a correct answer. Figure 10 represents the results of a single subject, who applied five different expressions for his subjective probabilities. Figure 11 gives the difficulty-by-label curves for all subjects using five labels. The curves are fairly parallel and the median values of the possibility functions (ranges of applicability for the expressions) are in good agreement with the probabilities estimated from the model.

In order to make these results more comparable to the ones depicted in Figure 10 I have computed the probabilities of successes for given ability levels and given difficulty levels conditioned on the labels applied. These "objective" probabilities are plotted against the median values of the possibility functions representing the meaning of the verbal expressions in Figure 12.

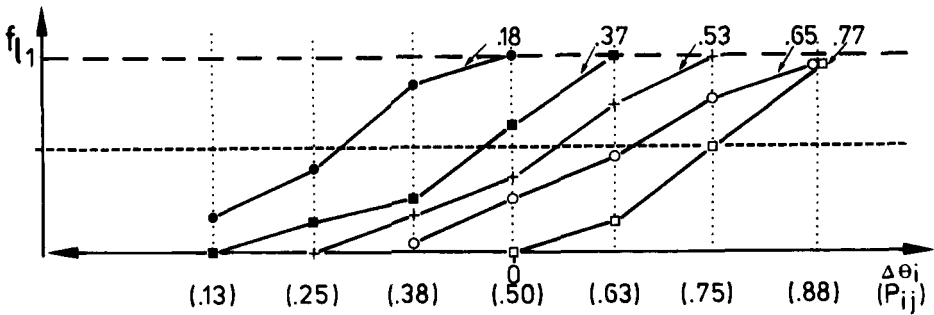


Figure 10. The cumulative frequencies for the verbal labels of one subject. The curves are drawn for items from seven categories of difficulty for this subject, which determine the individual probabilities of successes according to the Rasch model. The curves are labelled with the mean frequencies of the second experiment (see Figure 3). The subject would be perfectly calibrated if the labelling of the curves and the cutting points with the dashed line were the same. This subject is slightly underconfident for hard items but well calibrated for the easy ones.

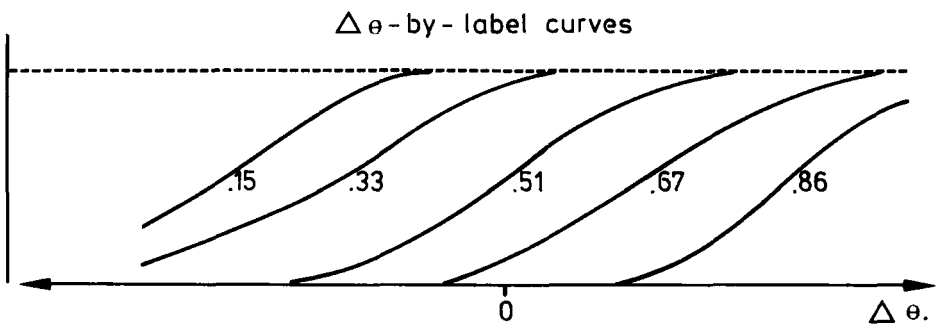


Figure 11. The average cumulative difficulty by label curves for all subjects applying five expressions for their subjective probabilities of successes

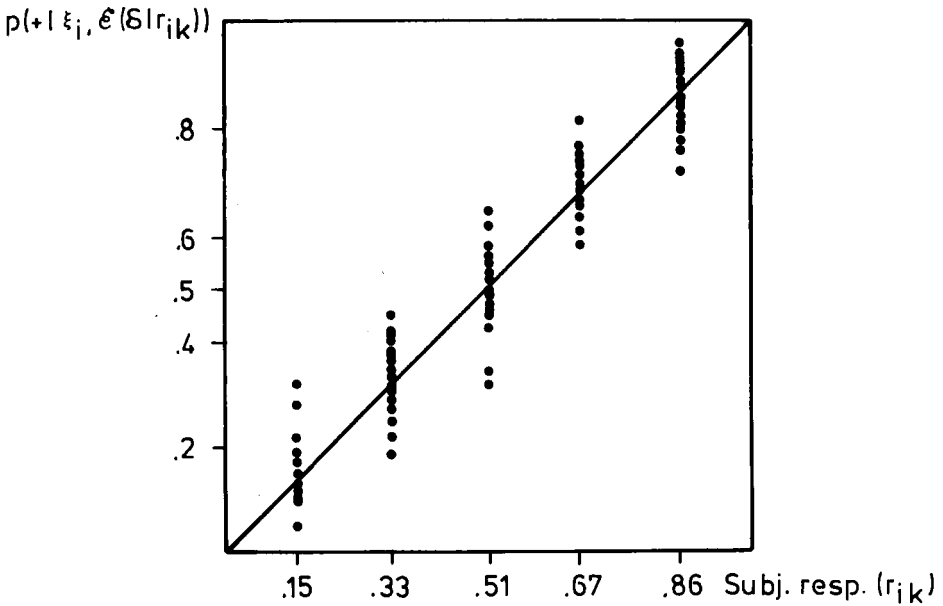


Figure 12. The probabilities of successes for given ability and difficulty levels conditioned on the applied verbal expressions plotted against the median values of the possibility functions for these verbal expressions.

Although there are marked variances, the results indicate that on the whole this group of subjects is well calibrated.¹

¹ SCHUETT (1981) reports that after sufficient training and optimal support by a computerized information processing system his subjects were able to estimate subjective probabilities, which did not violate SAVAGE's (1954) axioms. This result does not contradict the conclusions I have drawn from my experiments; that is, humans usually express their confidence verbally and they are able to perform by means of verbal arguments a task which is as complex as SAVAGE's system of axioms. Furthermore the results reported here indicate that by using this kind of information processing people are able to give veridical judgments without the amount of external support provided in SCHUETT's (1981) experiment.

Further Applications of this Model

As pointed out in the beginning I assume that any mode of processing information which imposes higher mental work load is bound to render any consecutive rating more biased than a mode imposing a lower mental work load. Examples for the negative effect of a high mental work load can be found in the studies done on conservatism in information processing and on the negligence of the regression effect in numerical prediction tasks. In order to find out if the application of verbal expressions in these tasks relieves the biasing effect, I have applied my model to two prototypical experiments from this area.

Conservatism

A typical example for the suboptimal information processing has been demonstrated in experiments on conservatism, where subjects tend to stick to their initial assumptions concerning the probabilities of events despite the fact that in the light of new information they should revise these assumptions. The optimal revision strategy for assumptions in this task is the application of Bayes' theorem, which therefore can be used as a normative standard for the subjects' performance. In a series of experiments PHILLIPS and EDWARDS (1966) have investigated this phenomenon with the result that in all cases conservatism occurred. This effect could be slightly reduced by relieving the memory load of the subjects and in one experiment by permitting them to answer verbally. I have replicated part of their experiment III with 30 subjects (undergraduates in economy). Subjects were shown two identical bins pictured side by side on a tv screen. They were told (A) contained 70% bricks and 30% balls and the other (B) 30% bricks and 70% balls. On each trial subjects pushed a button and thereby drew a sample of one - either a brick or a ball - from each of the bins. Subjects were asked to describe verbally the chance of getting a brick or a ball the next time from the bin. On a second screen a tally was kept of the numbers of bricks and balls drawn from each bin, together with the last verbal expression the subject had given for the chances. The results plus the ones from PHILLIPS and EDWARDS (1966, experiment III) are shown in Figure 13. These results clearly indicate that verbal responses together with a visible record induces near optimal, or, Bayesian performance in the subjects. In the light of this results one can assume that the so

Conservatism

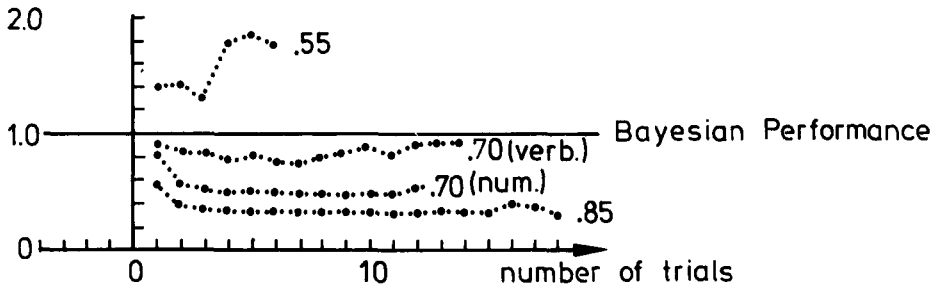


Figure 13. Comparison of the results from PHILLIPS and EDWARD's experiment III with the ones from our experiment with verbal expressions for chances (labelled ".70 (verb.)")

called conservatism in information processing is mainly due to a conservative handling of numbers and computation and is not necessarily due to a biased processing procedure.

Prediction

In their 1973-study KAHNEMANN and TVERSKY have tried to determine which kinds of heuristics and biases influence subjects' performance in predicting future events. They studied categorical as well as numerical predictions and found in all cases that subjects tended to neglect the base rates and exhibited mistaken intuitions about regression. For instance, they acted as if in all cases the correlation between the predictor variable and the criterion were perfect and as if the predictor variable were measured with perfect reliability. In order to determine if these mistaken intuitions are at least partially caused by the apparent difficulties subjects have in processing

numerical variables, I have undertaken the following experiment: subjects (30 undergraduate students in psychology who had not attended courses in statistics) were asked to predict the success of students in the university from their performance at highschool. Subjects were given a sample of 75 cases which includes the grades the students received at highschool and in the graduating exam at the university. The results were very similar to the ones obtained by KAHNEMAN and TVERSKY (1973) but when the subjects were asked how certain they were about the correctness of their predictions, they gave quite low subjective degrees of confidence. When probed further about the direction of their probable error of prediction 27 subjects indicated correctly that the true value would probably be closer to the mean performance than they had predicted. Only 3 subjects answered that they could not tell.

This result as well as those reported above seem to indicate that subjects are better able to take into account complex dependencies by means of verbal processing than they are if they are forced to process the same amount of information numerically. As GREGORY (1982) pointed out, numbers and computation form a more recent tool of mind than language and therefore the numerical information processing is less automatic.

A further aspect of the verbal superiority is illustrated by the outcome of an experiment from a different study. In this experiment² subjects (24 bank clerks responsible for foreign exchange) were asked to predict what the exchange rate between the US Dollar and the Deutschmark would be four weeks later. 12 Subjects had to give the predictions "in their own words as they would talk to a client", whereas the other twelve were asked to give numerical estimates in percentage of change. Both groups were asked to verbalize the steps they took in order to come up with the prediction. After the predictions of the first group had been calibrated with a technique similar to the one of our second experiment, they were compared to the one made by the numerical forecasting group. It turned out that the first group was more

² In a different context I have reported other aspects of the results of this experiment.

correct and more internally consistent. While this is interesting in itself, another point is more important: The slight difference in the instructions given caused marked differences in the way the subjects performed their task as revealed by verbalizing. The verbal prediction group used quantitative variables (e.g. the GNP increase in percent) as well as qualitative variables (e.g. the stability of the German government) for deriving their predictions, whereas the other group merely took into account those variables which were usually expressed numerically. From this it seems plausible to assume that the reason for the superiority in the verbal forecasting condition is the fact that the knowledge base on which these subjects relied was broader and allowed for more elaboration.

An alternative interpretation of those results is possible too: It might be that the thinking-aloud interfered specifically with the numerical reasoning.

General Conclusion

Although the studies reported here have originated from a perspective of applied psychology, aiming at better ways to implement vague knowledge into decision processes, there may be consequences of this kind of research for cognitive psychology in general. When analyzed by formal means it turns out that decision making and forecasting on the one hand and reasoning and problem solving on the other hand are structurally identical. Therefore it can be assumed that the dependency of cognitive processes on the underlying mental representations might be a quite general problem.

The experimental results presented here indicate that: first, the available procedures for human information processing are strongly dependent on the modes of mental representations, in which they are applied; and second, human subjects are more effective in reasoning with verbal expressions than with numerical expressions, even if the tasks performed rely on frequency information.

Reference Notes

1. personal communication

2. personal communication
(ZIMMER, note 3)
3. ZIMMER, A. A model for the interpretation of verbal predictions, manuscript, Stanford, 1983, (submitted for publication)

References

- BEGG, I. (1982) On the interpretation of syllogisms. *Journal of Verbal Learning & Verbal Behavior*, 21, 595-620
- DE FINETTI, B. (1937) La prévision: des logiques, ses sources subjective. *Annales de l'Institut Henri Poincaré*, 7, 1-68
- DUBOIS, D. & PRADE, H. (1980). *Fuzzy Sets and Systems: Theory and Application*. New York: Academic Press
- FERRELL, W.R. & MC GOEY, D.J. (1980). A model of calibration for subjective probabilities. *Organisational Behavior and Human Performance*, 26, 32-53
- FREKSA, C. (1981) *Linguistic Pattern Recognition*. Ph. D. thesis, UC BERKELEY
- FREKSA, C. (1982) Linguistic description of human judgments in expert systems and in "soft" sciences. In: L. TRONCALE (ed.) *Systems Science and Science*. Louisville, Ky.: Society for General Systems Research
- GREGORY, R.L. (1981) *Mind in Science*. Cambridge: Cambridge University Press
- HACKING, I. (1975) *The Emergence of Probability*. Cambridge: Cambridge University Press
- HERSH, H.M. & CARAMAZZA, A. (1976) A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology, General*, 105, 254-276
- HINTZMAN, D.L., NOZAWA, G. & IRMSCHER, M. (1982) Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 127-141
- KAHNEMAN, D. & TVERSKY, A. (1973) On the psychology of prediction. *Psychological Review*, 80, 237-251
- KAUFMANN, A. (1975) *Theory of Fuzzy Subsets*. Vol. 1. New York: Academic Press
- LICHTENSTEIN, S., FISCHHOFF, B. & PHILLIPS, L.D. (1982) Calibration of probabilities: The state of the art to 1980. In: KAHNEMAN, D., SLOVIC, P., TVERSKY, A. (eds.) *Judgment under uncertainty. Heuristics and Biases*. Cambridge: Cambridge University Press

- LINDLEY (1974) Discussion on the papers by Professor Tversky and Professor Suppes. *Journal of the Royal Statistical Society, Series B.* 36, 181-182
- NISBETT, R.E. & ROSS, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, N.J.: Prentice Hall
- NISBETT, R.E. & WILSON, T.D. (1977) Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259
- PHILLIPS, L. & EDWARDS, W. (1966) Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72, 346-354
- RASCH, G. (1966) An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57
- SAVAGE, L.J. (1954) *The Foundations of Statistics.* New York: Wiley
- SCHÜTT, K. P. (1981) *Wahrscheinlichkeitsschätzungen im Computer-Dialog.* Stuttgart: Poeschel Verlag
- SLOVIC, P., FISCHHOFF, B. & LICHTENSTEIN, S. (1979) Images of disaster: Perception and acceptance of risks from nuclear power. In: G.T. GOODMAN & W.D. ROYCE (eds.) *Energy Risk Management.* London: Academic Press
- SLOVIC, P. & TVERSKY, A. (1974) Who accepts Savage's axiom? *Behavioral Science*, 19, 368-373
- SUPPES, P. (1974) The measurement of belief. *Journal of the Royal Statistical Society, Series B*, 36, 160-175
- TVERSKY, A. (1974) Assessing uncertainty. *Journal of the Royal Statistical Society, Series B*, 36, 148-159
- WITTE, W. (1960) Struktur, Dynamik und Genese von Bezugssystemen. *Psychologische Beiträge*, 4, 218-252
- WITTGENSTEIN, L. (1969) *Philosophische Untersuchungen.* Frankfurt: Suhrkamp
- WRIGHT, G.N. & PHILLIPS, L.D. (1980) Cultural variation in probabilistic thinking: alternative ways of dealing with uncertainty. *International Journal of Psychology*, 15, 239-257
- ZADEH, L.A. (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3-38
- ZIMMER, A. (1980) Eine Formalisierung mnestisch stabilisierter Bezugssysteme auf der Grundlage von Toleranzmengen. In: A. THOMAS & R. BRACKHANE (eds.) *Wahrnehmen - Urteilen - Handeln.* Bern: Huber
- ZIMMER, A. (1982 a) What really is turquoise? A note on the evolution of color terms. *Psychological Research*, 44, 213-230
- ZIMMER, A. (1982 b) Some experiments concerning the fuzzy meaning of logical quantifiers. In: L. TRONCALE (ed.) *Systems Science and Science.* Louisville, K.: Society for General Systems Science

SUBJECTIVE EXPECTED UTILITY: A MODEL OF DECISION-MAKING¹⁾

BARUCH FISCHHOFF
BERNARD GOITEIN ZUR SHAPIRA

Decision Research
A Branch of Perceptronics
USA

A simple and comprehensive rule for making decisions is the following: List all feasible courses of action. For each action, enumerate all possible consequences. For each consequence, assess the attractiveness or aversiveness of its occurrence, as well as the probability that it will be incurred should the action be taken. Compute the expected worth of each consequence by multiplying its worth by its probability of occurrence. The expected worth of an action is the sum of the expected worth of all possible consequences. Once the calculations are completed, choose the action with the greatest expected worth.

Studying the prescriptive and descriptive validity of this model is the pre-occupation of researchers in a diffuse field known as behavioral decision theory (BDT). Some look at the model as a whole, others concentrate on its components (e.g., assessments of worth or probability), and still others concern themselves with special cases. The most general form is known as the subjective expected utility (SEU) model. In it, probabilities are treated as being subjective, while worth is expressed in *utility*, a generalized measure

¹⁾ Adapted from a paper by the same authors entitled "The Experienced Utility of Expected Utility Approaches." In: N.T. Feather (ed.), *Expectations and Actions*. Hillsdale, NJ: Erlbaum Associates, 1981.

of (un)desirability. The "S" is dropped if probabilities are considered to be objective (a distinction clarified below). The "U" becomes a "V" (for "value") if an absolute standard of worth, like dollars, is used. The E is dropped if probabilities are ignored. Finally, the various attributes of each consequence may or may not be considered explicitly. When they are, the rubric "Multiattribute" is invoked. This essay traces the history of SEU research, with particular emphasis on its varied intellectual roots and the insights and blind spots they have entailed.

One point of reference for the review will be the concurrent, but relatively independent, work done under the rubric of Expectancy-Valence (E-V), research [1]. Although they have used SEU models, E-V investigators have been less interested in understanding the general properties of the model than in using it to predict and abet human performance in specific applied settings, such as career counseling or employee management. The experiences and beliefs of E-V and BDT researchers provide interesting counterpoints. The great advantage of E-V studies is the realism of their settings; their great disadvantage lies in the complexity of those settings. Many of the conclusions of behavioral decision theory regarding the intricacies of decision behavior and its study were discovered and documented slowly and painstakingly in highly constrained experimental situations. Such subtle signals might never have been detected in the flesh. Conversely, field research of the E-V type reminds us of the key variables that should be presented in laboratory work.

Formal Models of Decision-Making

Attempts to describe how people make decisions took their initial marching orders from prescriptive models, telling how people *should* make decisions. The appeal of SEU models in this report is obvious. They reduce the universe of decisions to a common set of primitives (options, probabilities, utilities) about which one could hope to derive universal truths (e.g., place similar utilities on losses and gains of the same absolute magnitude) in due time. To some extent, this promise has been borne out by reality. Three hundred articles written yearly on behavioral decision theory probably produce much more cumulative knowledge than three thousand publications on personality [2]. A second enticement is the presence of an axiomatic basis for these

models. Any individual who subscribes to a small set of reasonable rules (e.g., transitivity of preferences) should behave according to the model. Prescriptive reasonableness is presumably a necessary condition for the descriptive validity of a theory of conscious behavior (i.e., whatever rule people are supposed to be following - according to one's theory - should make sense to them).

In mainstream American economics, the idea that people maximize V or U or $E-V$ or SEU or multiattribute SEU has the status of a metatheory; it goes without saying that such maximization is what people try and often manage to do. The goal of the theorist is not to test this assumption, but to divine just what it is that people are maximizing. With sufficient ingenuity, one always finds something that a particular decision-maker has maximized in a particular situation. Like other metatheories, SEU is not falsifiable and thus not a theory at all, but rather a rule in this game.

The story of behavioral decision theory has, however, been the growing realization that SEU often does not describe the decision-making process either as it is designed or executed. The dramatic tension has been provided by SEU 's remarkable ability to hang on despite mounting doubts about its descriptive competence. The climax of this yet incomplete tale seems to have something to do with realizing that a clearly erroneous model still may serve some useful role. The denouncement (or perhaps sequel) may involve elaborating and circumscribing that role. The retrospective look which follows suggests that the secrecy of its longevity lies in the fact that the model (a) expresses some fundamental truth in some situations, (b) has a remarkable ability to be close to the mark in situations where its underlying logic is clearly inappropriate, and (c) has proponents of heroic tenacity.

Model Fitting

The typical way to test the descriptive validity of the SEU model is to see if it can "explain" past decisions. A set of similar decisions is collected and the researcher sets about identifying a set of predictors (probabilities and utilities) that, when combined according to the SEU rule, will account for those decisions.

Underlying this approach is a fairly reasonable research strategy. It is extraordinarily difficult to study simultaneously the form of the decision rule that people use and the substantive considerations incorporated in them; therefore, assume the truth of a general model and concentrate on identifying the inputs that will make it work. Specifically, assume that people use the SEU model and find what their probabilities and utilities are. With such a strategy, it is hard to lose in the short run. The clever researcher can usually find some sets of values in a particular situation that will give good enough predictions or postdictions to maintain faith in the model. Moreover, any descriptive failure by an SEU model can be explained away as a failure in measurement rather than the use of the wrong combination rule. The wrong consequences were included; their worth was assessed in an unreliable or confusing manner; uncertainties were ignored or elicited improperly and so on.

With the complementary model-fitting strategy one must assume that one knows the probabilities and utilities to which people attend and then must experiment with different combination rules until one is found that works. The price paid for strict adherence to either strategy is in the long run. As the number of predictors (different variables, measurement techniques, data transformations, etc.) increases, the model becomes increasingly vague and imprecise. Given a sufficiently large set of probabilities and utilities, one can, of course, devise a rule "predicting" past decisions to any desired level of proficiency. In regression terms, by expanding the set of independent variables one can always find a set of predictors (or even one predictor) with any desired correlation with the dependent variable. The price one pays for overfitting is shrinkage, failure of the rule to "work" on a new sample of cases. With well defined problems, the predictor/sample ratio allows one to "correct for shrinkage" and estimate the predictive validity of an SEU model for future decisions. Such corrections are impossible when it is unclear just how many variants of the model have been tried and discarded before the best-fitting one was identified. Thus, when an SEU model "works" on a set of data, it is difficult to know that that proves. Have we proven that the decision-makers we are studying are using SEU, and here are their weights? Or has our fishing expedition finally stumbled upon an SEU-like model that mimics their actual behavior?

Simple Models or Simple Process?

Growing unease with the elusiveness of the proper SEU model was compounded by results emerging from the study of clinical judgment, another intellectual tradition upon which behavioral decision theory draws. Clinical judgment is exercised by a radiologist who sorts x rays of ulcers into "benign" and "malignant", by a personnel officer who chooses the best applicants from a set of candidates, by a crisis center counselor who decides which callers threatening suicide are serious. In each of these examples, the diagnosis involves making a decision on the basis of a set of cues or attributes. When, as in these examples, the decision is repetitive and all cases can be characterized by the same cues, it is possible to model the judges' decision-making policy statistically. Following the logic of the SEU modeling described in the previous section, a set of decisions are gathered and multiple regression is used to determine the cue-weighting scheme that best accounts for the decisions actually made by the judge. Two decades of such *policy-capturing* studies persistently produced a conclusion: Simple linear models, using a weighted sum of the cues, did an excellent job of postdicting judges' decisions, despite the judges' claim that they were using much more complicated strategies [3-5]. A commonly asserted form of complexity is called "configural" judgment, in which the diagnostic meaning of one cue depends on the meaning of other cues (e.g., "that tone of voice makes me think 'not suicidal' unless the call comes at midday").

Two reasons for this contrast have emerged, each with negative implications for the descriptive validity of SEU. One involves the growing realization that combining large amounts of information in one's head overwhelms the computational capacity of anyone but an *idiot savant* [6,7]. Thus, a judge trying to implement a complex strategy simply would not be able to do so with great consistency. Since the SEU model is such a rule, these results suggest that people could not use it even if they tried.

The second realization that emerges from clinical judgment research is that simple linear models are extraordinarily powerful predictors. If one can identify and measure the attributes that decision-makers consider, one can mimic their decisions to a large degree with simple models bearing no resem-

blance to their cognitive processes. That is, one can misspecify weights and even combination rules and still do a pretty good job of predicting decisions under very general conditions [8]. Thus, whatever people are doing will look the application of an additive linear model (such as SEU). In Hoffman's [9] terms, such models are *paramorphic* in that they can replicate the input-output relationships of phenomena without any guarantee of fidelity to the underlying processes.

The response of the research community to these problems was a parting of ways between two camps which might be described as involving cognitive applied and applied cognitive psychologists [10,11]. The latter were interested in how people think in the general applied context of decision-making; their progress is described in subsequent sections.

The former were interested in solving specific applied problems involving cognitive decision-making. For them, the predictive power of these models made them a highly useful tool. Often one does not care how decisions are made, as long as one can predict their outcome. If it works, even an admittedly paramorphic model may be good enough for designing an effective marketing campaign or remuneration scheme. Indeed such a model might be used to replace clinical judges with formulas modeling their behavior. Called "bootstrapping", this approach has decision-makers identify the variables upon which they base their decisions and then "captures" the policy they use in a set of trial cases. That policy, as embodied in a formula, is then used in place of the judges. Even if it does not embody their thought processes, a formula may mimic their decisions with greater reliability (and hence validity) than the judges themselves. Formulas never have off-days or suffer from fatigue or distraction by irrelevant cues [5]. Clearly, people know more than is included in such formulas. However, there is no evidence to date indicating that they can convert this sensitivity into superior predictions.

Empirically discovering an analytical result by Wilks [12], Dawes and Corrigan [13] showed that considerable predictive success is possible with even minimal modeling. All one has to do is identify the variables (or attributes) to which a decision-maker attends and decide whether they are related positively or negatively to the decision criterion (i.e., whether they are good

or bad signs). If these variables are expressed in standard units, they can be given unit weights (+1 or -1, as appropriate). Such a unit-weighted model will, under very general conditions, predict decisions as well as a weighted-sum model using regression weights and is analytically quite similar.

Thus, a simple substantive theory indicating what variables people care about when making decisions may be all one needs to make pretty good predictions of their behavior. If they like more of good things and less of bad things, just count up the number of good things an option leads to and subtract the number of bad things and you have a good idea of how favorably it will be viewed. The predictive power of such a simple model provides a baseline against which more sophisticated models must be compared.

Obviously, some goods (positive attributes) are more important than others. Therefore, a model using importance weights should, in principle, predict better than one using unit weights. Similarly, goods obtained with high probability should be valued more than those obtained with low probability. Therefore, a model using probability weights in principle should predict better than one using unit weights. However, any unreliability or misspecification of those weights reduces their contribution very rapidly. In the extreme, models using poorly conceived or executed weighting schemes may succeed in spite of rather than because of their differences from the unit-weight model.

Divide and Conquer

Feeling that the SEU model was impregnable and impenetrable when dealt with as a whole, the applied cognitive psychologists among behavioral decision theorists have largely turned to an examination of the model's parts and their validity. The following sections describe the answers found by asking: (1) Do people accept the axioms upon which the model is based? (2) Are they capable of providing the inputs which it requires? (3) In simplified situations, which relieve their computational load, do they exhibit SEU-like behavior?

Prescriptive Validity

One great attraction of SEU approaches is that they represent reasonable decision rules, ways in which people would want to make decisions. The SEU model states that options become more attractive as their positive consequences become more appealing or more likely, or as their less appealing consequences become less likely. Since the expected utilities of various possible consequences are added, one consequence's low or negative utility can, in principle, be compensated for by sufficiently high utility on others. If people endorse these rules, then the task of descriptive research is just studying how (well) they are carried out.

When might people reject these rules?

(a) When probabilities and values are not independent. One form of dependence is when the probability of receiving a consequence affects the value ascribed to it. For example, people will sometimes forego an increment in probability of success in order to be challenged by a situation, or to have some suspense associated with its outcome, or to learn something from how it turns out [14]. One might prefer a class in which the top 80 % receive A's to one in which A's are automatically given to everyone. Assigning intrinsic values to probabilities of success violates the SEU model's axioms. However, it is possible to redefine the consequences so as to save the model. Thus, the "cinch A" and the "earned A" might be considered different consequences whose utilities differ enough that a 1.0 chance of the former is less attractive than a 0.8 chance of the latter.

Influence in the opposite direction is also possible. Probabilities may be influenced by utilities, as when optimists exaggerate the probability of good things happening. For example, Weinstein [15] has found that people typically believe that they are more likely than the average individual to have good things happen and less likely to suffer various maladies. It is more difficult to rescue them from these deviations.

(b) When options are evaluated by noncompensatory criteria. Two possible noncompensatory rules are the conjunctive and disjunctive. By the conjunctive

rule, an option has to score fairly high on each consequence to be considered. For example, a vacation option must be reasonably priced, available when needed, suitably sunny and fairly quiet. If an option fails to pass muster on any one of these attributes, its rating on the others would be immaterial, e.g., no amount of sun will compensate for a lot of noise. According to the disjunctive rule, an option that is adequate on any one attribute is acceptable. For example, an investment opportunity might be chosen if it were good enough as a speculation, tax shelter, or hedge against inflation, no matter how badly it rated on the other dimensions. Investment portfolios often include varied items chosen by the same disjunctive rule. Compound strategies are also possible, such as using a disjunctive or conjunctive rule to reduce a large set of options to a smaller one to which a compensatory rule is then applied [16]. Although any reliance on noncompensatory rules would threaten the prescriptive validity of SEU models, Einhorn [17] found linear models to be so powerful that they still provided relatively good predictions of the behavior of people clearly using noncompensatory models. Choice theory [18], however, offers a variety of other noncompensatory decision rules whose use has been observed in one situation or another. Attempts to predict the usage of such strategies have a distinctly *ad hoc* character.

(c) When consequences are not evaluated independently. Literal usage of an SEU model requires one to evaluate each consequence of each option separately and then combine those assessments. In some situations, however, the decision-maker might want to evaluate a particular consequence differently depending on the value of other consequences. For example, one might like either a slightly uncomfortable chair or a moderate level of ambient noise when deciding where to sit in a lecture, knowing that either will help one stay awake. However, a seat that is both hard and located in a noisy area will be undesirable since it diverts too much attention. Using the language of judgment research, it reflects configural judgment. In other words, worth is an interactive function of the attributes. Such situations can, in principle, be handled by evaluating attributes in pairs or triplets rather than separately (i.e., one could evaluate a soft seat in a noisy area, a hard seat in a noisy area, a soft seat in a quiet area, and hard seat in a quiet area). With all but the simplest of attribute-attribute interactions, however, this would prove quite laborious.

(d) When options are not evaluated simultaneously. Simon's [7] notion of *satisficing* grew out of observing the predictive failures of the SEU-like models of classical economics. Satisficers look for any decision alternative that is "good enough". Their search terminates when a satisfactory option has been identified and evaluated. This option may, however, be inferior (in an SEU sense) to other options that have not even been considered, because the serial evaluation process was truncated prematurely.

(e) When people explicitly reject the SEU model's axioms. One way to see if people accept the axioms is to ask them. In 1969, Tversky [19] found that although people like to be transitive in their choices (thus fulfilling one axiom), one can design situations in which they are both intransitive and unable to resolve the inconsistency in their choices. He speculated that marketing approaches may be designed to exploit such difficulties. The Allais and Ellsberg paradoxes [20,21] are two demonstrations of people rejecting Savage's independence axiom, according to which preferences between alternatives should be independent of any consequences they have in common. Slovic and Tversky [22] presented subjects with arguments for and against this axiom and found persistent violations, even after the axiom was presented in a clear and presumably compelling fashion. The problem with such demonstrations is that they do not indicate the prevalence and distribution of violations. Are they just concocted curiosities? Or do they represent modal behavior in some important realms?

Kahneman and Tversky's [23] "prospect theory" attempts to lend system to violations of the SEU axioms by explaining them as results of consistent behavioral tendencies. One such tendency, the "certainty effect", causes consequences that are obtained with certainty to be given more weight than uncertain consequences. The second, labeled the "reference effect", leads people to evaluate alternatives relative to a reference point which may correspond to their present status, to some expectation about future status, or to some adaptation level [24]. By altering the reference point, alternative versions of the same decision problem that are formally equivalent in terms of the SEU model may elicit different preferences. Once a decision problem has been posed in a particular way, Kahneman and Tversky's theory makes powerful predictions about responses to it, predictions that are often con-

trary to predictions of the SEU model. It needs, however, to be complemented by a substantive theory regarding the way in which decision questions are posed by nature and interpreted by decision-makers.

Availability of Inputs

People who wish to engage in SEU-like behavior must provide a list of all feasible options and possible consequences, along with assessments of their probability and utility. One need not perform these tasks well in order to attempt SEU-like behavior. Poor performance will, however, lead to suboptimal decisions, relative to what could be obtained by the best possible usage of the SEU model. Not making the best decisions need not mean that people are not trying to use the SEU model. Nonetheless, acute inability to provide the inputs required to use the SEU model well must cast some doubts on its popularity, for why would people persist in doing something they do quite poorly? Whether it is reasonable to expect people to persist at suboptimal SEU behavior, rather than opting for some simpler rule, requires consideration of three questions: (a) Just how good (or bad) are they? (b) How likely are they to learn about their mistakes? (c) What are their options? If people could not perform these tasks, realize their limitations, have viable options, the SEU model would be a less likely candidate for how people try to make decisions.

What else could people do, besides try to list all courses of action open to them and assess their expected consequences? One option is try not to think at all, but rely on nonanalytical decision rules, such as "this is (most like) what I've always done" or "this is what my (most expert) friends tell me to do" or "this is what everyone else is doing". Another option is refuse to make a (final) decision, preferring to muddle through by trial and error, making small incremental decisions that afford an opportunity to change courses if things are not going well [25]. A third option is to recruit a number of people to think about issues too complex and poorly understood to analyze alone and rely on whatever conclusion these chaotic group processes produce. Such reliance on group rather than individual "rationality" is common in work situations [26,27], with perhaps the best known example being the emergence of group production norms in defiance of piecework pay schemes.

Problem Structuring

Like other analytic approaches, SEU begins with a structuring of the problem, namely listing all relevant options and consequences. The obvious performance standard here is completeness. A modest amount of systematic and anecdotal evidence suggests that people have difficulty independently producing complete problem representations. When asked to judge completeness of representations generated by themselves or others, however, they often seem to be very insensitive to these inadequacies. What is out of sight effectively is out of mind [28,29]. Thus in the short run at least, people may not realize this limit on their analytical decision-making ability. In the long run, omitting options and consequences should lead to poor decisions. However, life seldom sends us large enough batches of unambiguous signals to make it clear just how good our decision-making is and where the problems lie [30,31].

Assessing Probabilities

A cornerstone of SEU thinking is that we live in an uncertain world. According to the Bayesian or subjectivist position that uncertainty reflects the limits of our understanding. From this position, all statements are qualified implicitly or explicitly by our degree of belief in them. Degrees of belief are expressed numerically in subjective probabilities, i.e., expressions with the form "My personal probability that statement A is true is .XX.". Such probabilities are entirely in the eye of the beholder and are not properties of the world. Two observers of a situation in principle can assign different probabilities of being true to the same statement about it, either because they interpret what they see differently or have different background information.

In fact, subjectivists would argue that the whole notion of "objective probability" is meaningless. Assigning any probability requires some interpretation of the observed situation and that act of interpretation imputes a subjective element. What are commonly thought of as objective probabilities merely refer to situations in which there is consensus among reasonable observers about how to interpret existing evidence. The probability of a fair coin falling on "heads" on its next flip is commonly held to be 0.5, but

that is only because observers agree on how to evaluate past performance and relate it to subsequent performance (e.g., what constitutes an adequate series of independent, equivalent trials). Bayesians do not talk about *estimating* probabilities which implies that there is something "out there" in the world to be appraised, but of *assessing* probabilities, signifying the evaluation of an internal state.

Assessing people's performances in assessing probabilities poses some problems. Individual probabilities can almost never be evaluated (except in the case where one says "There is *absolutely* no chance of this statement being wrong" regarding a statement that is, in fact, wrong). Sets of subjective probabilities, however, may be evaluated according to two criteria. One is consistency with the laws of probability or (internal) coherence [32]. For example, the probabilities assigned to a statement and its complement should total 1.0. The second is "calibration", applicable when the truth of uncertain statements eventually becomes known. The well calibrated probability assessor should tend to have more true statements associated with high than with low probabilities. Specifically, XX % of the statements assigned probability .XX should be true. That is, if I am well calibrated, 70 % of the time when I state "there is a 0.7 chance of this statement being true", it should be true.

The earliest studies of probability assessment used repetitive series of events (like drawing red or blue balls from an urn) and defined probability as relative frequency of occurrence. Research showed that people were quite good at appraising such frequentistic probabilities [33].

More recently, the focus of research has turned to how people assign probabilities to statements, including unique events (e.g., Carter will be re-elected), where relative frequency has no meaning. Here, performance has been less than outstanding. Studies using a variety of tasks, response modes, and subjects have shown that probability assessments tend to be calibrated poorly [34]. The most common error is overconfidence, i.e., people think that they know more than is actually the case [35]. Where observed, this bias seems to be so robust that people are willing to engage in highly disadvantageous bets based on their confidence judgments [28].

Various forms of incoherence have also been observed, although without any clear indication of their prevalence. For example, Wyer [36] found that people tend to exaggerate the probability of the conjunction of two events, relative to the probabilities assigned to the two individually. Kahneman and Tversky [37] found that rather than combining base-rate (i.e., background) information with information regarding a particular case, people tend to ignore base-rate probabilities. Slovic, Fischhoff, and Lichtenstein [38] demonstrated situations in which the judged probability of a compound event was larger than the probabilities of its constituent events. A more fundamental level of confusion might be found in the folklore of survey research which holds that one cannot even ask for numerical probability assessments from a random sample of the population without getting disturbingly high rates (i.e., over 20 %) of nonresponse [39]. On the other hand, people seem to have a good appreciation of the meaning of probabilistic weather forecasts [40].

Perhaps the most disturbing aspect of these results for a proponent of SEU as a descriptive model is not the presence of biases, but the thought processes that seem to underlie both erroneous and accurate probability assessments. Tversky and Kahneman [41] have argued that many probabilistic judgments are produced by using rules of thumb or heuristics whose internal logic bears little resemblance to the rules of probability. These are generally effective ways of coping with an uncertain environment that both deny its probabilistic character and preclude learning about some basic phenomena (like regression to the mean). Often they lead to substantially biased judgments and to decisions not in the individual's best interest. To the extent that these heuristics capture people's thought processes, they argue (a) that we are not efficient probabilistic functionalists, and (b) that even when people do incorporate uncertainty in their decision-making, they do so in terms quite different from those of the probability calculus of the SEU models.

Assessing Values

The study of value judgments and their validity has languished relative to the study of probability judgments. The reason for this apparent disinterest is probably the absence of an acceptable criterion by which to evaluate value

judgments. Such judgments would seem to be the last redoubt of unaided intuition. Who knows better than a respondent what he or she prefers?

Recent research has, however, revived and elaborated a lesson long known to attitude researchers. Subtle changes in how value questions are phrased and responses are elicited can have marked effects on the preferences people express. This lability in value judgments seems to have important implications for both how values are conceptualized and how they are studied.

An article of faith among students of value, choice, and attitude judgments is that people have reasonably well defined opinions regarding the desirability of different events. Although these opinions may not be formulated intuitively in numerical (or even verbal) form, careful questioning can elicit judgments representing people's underlying values. From this stance, elicitation procedures are neutral tools, bias-free channels which translate subjective feelings into scientifically usable expressions. They impose no views on respondents, beyond focussing attention on those value issues of interest to the investigator.

What happens in cases where people do not know, or have difficulty appraising, what they want? Under such circumstances, elicitation procedures used in the research may become major forces in shaping the values expressed, or apparently expressed, in the judgments they elicit. Experimenters can induce random error (by confusing the respondent), systematic error (by hinting at what the "correct" response is), or unduly extreme judgments (by suggesting clarity and coherence of opinion that are not warranted). In such cases, the method becomes the message, a reflection of the researcher's tools rather than the respondent's wishes.

People are most likely to have clear preferences regarding issues that are familiar, simple, and experienced directly. Each of these properties is associated with opportunities for trial-and-error learning, such as might be summarized in readily applicable rules, that provide stereotypic, easily justifiable responses to future questions of values. When adopted by individuals, they may be seen as habits; when adopted by groups, they constitute traditions.

The power of these rules of thumb for assessing values comes from their development and application to simple and repetitive problems. Their viability becomes suspect when the issues are unfamiliar and complex, the old intuitions irrelevant, the old rules untested. Unfortunately, these are precisely those situations in which values need studying. In them, however, we may never have considered the implications of the values and beliefs acquired in simpler settings. As a result, we may have no articulated preferences. In some fundamental sense, our values may be incoherent, not thought through. In thinking about what are acceptable levels of risk, for example, we may be unfamiliar with the terms in which issues are formulated (e.g., social discount rates, minuscule probabilities, or megadeaths). We may have contradictory values (e.g., a strong aversion to catastrophic losses of life and a realization that we are not more moved by a plane crash with 500 fatalities than one with 300). We may occupy different roles in life (parents, workers, children) which produce clearcut, but inconsistent values. We may vacillate between compatible, but strongly held, positions (e.g., freedom of speech is inviolate, but should be denied to authoritarian movements). We may not even know how to begin thinking about some issues (e.g., the appropriate tradeoff between the opportunity to dye one's hair and a vague, minute increase in the probability of cancer 20 years from now). Our views may undergo changes over time (say, as we near the hour of decision or the consequence itself) and we may not know which view should form the bias of our decision.

Such inarticulated preferences are hardly compatible with the sort of rigorous systematic thinking required by SEU decision-making. Furthermore, they give pause for thought about just what it is subjects are giving us when we elicit the values needed to test the conformity of their behavior with the SEU model.

Listing a few specific effects may indicate the power an elicitor may deliberately, or inadvertently, wield in shaping, distorting, or even creating expressed preferences. One such effect mentioned earlier was the ease with which one can alter the desirability of outcomes by shifting the reference point in the light of which they are evaluated [37]. The same safety program might look quite different when conceptualized in terms of "lives saved" or "lives lost", with the respective reference points of the current situation

or an ideal one. A familiar effect from the study of motivation is that as one gets closer to an event with mixed consequences, the aversiveness of its negative aspects increases more rapidly than the attractiveness of its positive aspects, making it appear, on the whole, less desirable than it did from a distance [42]. People may have opposite orders of preference for gambles when asked which they prefer (which focusses their attention on how likely they are to win) and when asked how much they would pay to play each (which highlights the amount to win) [43]. People may prefer to take a chance at losing a large sum of money rather than absorb a smaller loss, but change their mind when the "sure loss" is relabeled as an "insurance premium" [44]. A relatively unimportant attribute may become the decisive factor in choosing between a set of options if they are presented in such a way that the attribute affords the easiest comparison between them [45] .

Three important features of these shifting judgments are (a) people are typically unaware of the potency of such shifts in their perspective, (b) they often have no guidelines as to which perspective is the appropriate one, and (c) even when there are guidelines, people may not want to give up their own inconsistency [44].

Elicitors must decide at some point whether they have captured their respondent's values adequately. The usual criteria are reliability and internal consistency (e.g., transitivity). However, where the task is poorly understood because the complexity or unfamiliarity, consistency of response within a given elicitation mode may only reflect the power of that mode to impose a particular perspective. Greater insight into values may come from posing diverse questions in the hope of eliciting inconsistent responses, in order to see how coherent the underlying values are and what changes make a difference.

In this light, the right way to ask a question depends on one's purpose in asking. If one is interested in how people express their values in the real world, one question may be enough. That world often asks only one question (e.g., in a ballot measure). A careful analysis of how an issue is posed may allow one to identify that question and predict responses accurately. If one is interested in what people really feel about a value issue, there may

be no substitute for an interactive elicitation procedure, one that acknowledges the elicitor's role in helping the respondent to create and enunciate values. That help would include a conceptual analysis of the problem and of the personal, social, ethical perspectives from which the respondent might wish to view it.

Decision Contexts

Although behavioral decision theorists have not developed a theory of decision situations to complement their evolving theory of the individual decision-maker upon occasion they have attempted to replicate their laboratory studies in the real world. These experiments at roughing it have been prompted by (a) the feeling that decision theory like other areas in applied cognitive psychology should work in the world; (b) the refusal of economists to accept (or even look at) laboratory data; (c) the hope of being able to elaborate their theories by embedding them in a broader context; and (d) continued interaction with decision analysts, those purveyors of SEU as a normative guide to decision-making who are also particularly attuned to the subjective quality of judgments of fact and value.

One obvious step in the direction of realism is to use real rather than hypothetical stakes in studies of decision-making. A second is to use experts rather than naive subjects. A third is using more realistic stimuli. Although available evidence is modest, none of these manipulations has provided sufficiently dramatic results to cast serious doubt on most of the research conclusions cited above [35,46].

None of these manipulations, however, leads to the study of actual judgments *in situ*. Probably the most elaborate excursion of behavioral decision theory into the real world has been a study by Kunreuther et al. [47], designed to understand the failure of people to buy federally subsidized flood insurance. That insurance is offered at highly attractive rates, where "highly attractive" was determined by planners who assumed that residents had accurate risk and cost data and combined them according to SEU principles. A national survey of flood-prone areas discovered, however, that residents' judgments of risks and costs were highly divergent from those of the experts. Often these

misjudgments seemed to have the same roots as the judgmental biases observed in laboratory studies. Obviously, even if residents were SEU decision-makers, they could not be expected to behave in the predicted way (i.e., purchase insurance) if they did not share the planners' information base. The design of the Kunreuther et al. study also allowed an appraisal of how favorable insurance should be to residents given what they thought to be the facts of the matter and assuming that they were SEU decision-makers. A straightforward derivation translated attractiveness of information to an index R , with higher R values being associated with more attractive insurance situations. Comparison of individuals who had and had not purchased insurance revealed little difference in the two distributions of R values. Thus, whatever guided insurance purchase decisions, it was not an SEU analysis based on either an "objective" (experts') or "subjective" (residents') view of the situation.

What did guide decisions? Kunreuther's best guess was that people relied on a series of informal decision rules such as the answers to the questions: Are my neighbors and relatives buying it? What does my insurance agent say? Do I have the capacity to worry about one more thing?

The application of expectancy value theory to predicting workers' effort and performance on their jobs represents a prolonged extension of SEU theory into the real world. Reviewing this effort, Schwab, Olian-Gottlieb, and Heneman [48] found that on the average only 9 % of the variance in effort or performance is explained by E-V models. Although their review considered only between-subject applications of this within-person theory of choice (thereby increasing the amount of error variance), analogous within-person studies rarely explain more than 25 % of the variance. Furthermore, with one exception, the few factors that Schwab, Olian-Gottlieb, and Heneman identify as contributing to better prediction are either omitted from SEU theory or are antithetical to it. They conclude: "there is a nagging suspicion that expectancy theory overintellectualizes the cognitive processes people go through when choosing alternative actions" [48, P. 146]. More explicitly pessimistic, Staw [49] suggests that ordinary choice behavior is better modeled as "impulse buying" decisions of consumers than on the lines of SEU theory.

Perhaps as revealing as the context to which laboratory results have been

extended are the contexts that have been avoided. Researchers' assumptions about situations to which their results will not generalize provide an implicit partial theory of decision-making environments. One such limiting feature of most behavioral decision theory experiments is that subjects are rewarded for responding as honestly as they can. However, there obviously are situations where it pays to lie (e.g., when the boss asks about work environment) and they may engender rather different behavior. A second limiting feature is the absence of time and emotional pressure that would create the sort of "hot cognition" studied by Janis and Mann [50]. A third feature is isolation of the individual from social interactions, which may both foster individualistic behavior and eliminate a common decision-making aid - talking to others. A fourth characteristic of most studies is the preclusion of a no-choice or procrastination option [51]. Each of these characteristics might represent an environmental variable restricting the validity of behavioral decision theory results or perturbation as minor as the shift from hypothetical to real stakes or from naive to expert subjects.

Conclusion

The story of SEU research has in some sense been a tale of deadends and hard-earned lessons. An enormous amount of effort was devoted to "capturing" decision strategies before it was realized that the power of linear models virtually precluded learning what weights individuals were giving to different attributes or if they were using an SEU decision rule at all. Many studies of how people estimate relative frequencies were conducted before researchers realized that such tasks failed to capture the sort of uncertainty found in most decision situations. For years, researchers derived satisfaction from the elicitation of consistent, reliable value judgments, before beginning to worry that such orderliness was a product of their methods.

Having disabused ourselves of these particular illusions, where do we go from here, in addition to having better respect for the power of our methods to produce tidy, but misleading results? Substantively, we should begin to develop an understanding of decision environments that will enable us to understand how general laboratory results are and how far-reaching the effects they produce. One key component of this understanding will be an

error theory explicating the consequences of poor performance in structuring decisions and in assessing and combining probabilities and values. The sensitivity of our decisions to such deviations from optimality will clarify which principles of decision-making need to be and can be learned from experience. Theoretically, we need to go beyond simply trying to falsify SEU theory, which, if only because of the power of linear models, will almost always provide at least mediocre predictions. Instead, we need to move on to sophisticated falsification [52], finding theories that do what SEU does and at least a little more. The shape of those theories might be new calculi, such as Prospect Theory, using different primitives and combination rules, or process-tracing models which attempt to treat not only input-output relationships, but also the intervening thought processes [16,53], or contingency-based models that preserve some of the SEU logic but also incorporate a theory of the environment and the way in which it is sequentially decomposed [54].

References

- [1] Feather, N.T. (ed.): 1981. *Expectations and Actions*. Hillsdale, NJ: Erlbaum
- [2] Goldberg, L.R.: 1974. Objective diagnostic tests and measures. In: *Annual Review of Psychology* 25, 343-366
- [3] Goldberg, L.R.: 1968. Simple models or simple processes? Some research on clinical judgements. In: *American Psychologist* 23, 483-496
- [4] Goldberg, L.R.: 1970. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. In: *Psychological Bulletin* 73, 422-432
- [5] Slovic, P. & Lichtenstein, S.: 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgement. In: *Organizational Behavior and Human Performance* 6, 458-479
- [6] Miller, G.A.: 1956. The magical number seven, plus or minus two: Some limits of our capacity for processing information. In: *Psychological Review* 63, 81-97
- [7] Simon, H.A.: 1957. *Models of man*. New York: Wiley
- [8] Dawes, R.M.: 1979. The robust beauty of improper linear models. In: *American Psychologist* 34, 571-582

- [9] Hoffmann, P.J.: 1960. The paramorphic representation of clinical judgement. In: *Psychological Bulletin* 47, 116-131
- [10] Baddeley, A.D.: 1977. Applied cognitive and cognitive applied psychology: The case of face recognition. Paper presented at Uppsala Conference on Memory.
- [11] Wright, P.: 1978. Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. In: *Instructional Science* 7, 249-312
- [12] Wilks, S.S.: 1938. Weighting systems for linear functions of correlated variables when there is no dependent variable. In: *Psychometrika* 3, 23-40
- [13] Dawes, R.M. & Corrigan, B.: 1974. Linear models in decision making. In: *Psychological Bulletin* 81, 2, 95-106
- [14] Atkinson, J. & Feather, N. (eds.): 1966. A theory of achievement motivation. New York: Wiley
- [15] Weinstein, N.D.: 1980. Unrealistic optimism about future life events. In: *Journal of Personality and Social Psychology* 39, 806-820
- [16] Svenson, O.: 1979. Process descriptions of decision making. In: *Organizational Behavior and Human Performance* 23, 86-112
- [17] Einhorn, H.J.: 1971. Use of nonlinear, noncompensatory models as a function of task and amount of information. In: *Organizational Behavior and Human Performance* 6, 1-27
- [18] Coombs, C.H./Dawes, R.M. & Tversky, A.: 1970. Mathematical psychology: An elementary introduction. Englewood Cliffs, NJ: Prentice-Hall
- [19] Tversky, A.: 1969. Intransitivity of preferences. In: *Psychological Review* 76, 31-48
- [20] Edwards, W.: 1954. The theory of decision making. In: *Psychological Bulletin* 51, 380-417
- [21] Edwards, W.: 1961. Behavioral decision theory. In: *Annual Review of Psychology* 12, 473-498
- [22] Slovic, P. & Tversky, A.: 1974. Who accepts Savage's axiom? In: *Behavioral Science* 10, 368-373
- [23] Kahneman, D. & Tversky, A.: 1979. Prospect theory. In: *Econometrica* 47, 263-291
- [24] Helson, H.: 1959. Adaptation level theory. In: S. Koch (ed.), *Psychology: A study of a science*, Vol. 1. New York: McGraw-Hill
- [25] March, J.G.: 1978. Bounded rationality, ambiguity and the engineering of choice. In: *Bell Journal of Economics* 9, 587-608

- [26] Salancik, G. & Pfeffer, J.: 1978. A social information processing approach to job attitudes and task design. In: *Administrative Science Quarterly* 23, 224-253
- [27] White, S.E. & Mitchell, T.R.: 1979. Job enrichment versus social cues: A comparison and competitive test. In: *Journal of Applied Psychology* 64, 1-9
- [28] Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1977. Knowing with certainty: The appropriateness of extreme confidence. In: *Journal of Experimental Psychology: Human Perception and Performance* 3, 552-564
- [29] Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. In: *Journal of Experimental Psychology: Human Perception and Performance* 4, 342-355
- [30] Fischhoff, B.: 1975. Hindsight $\neq$ foresight: The effect of outcome knowledge on judgement uncertainty. In: *Journal of Experimental Psychology: Human Perception and Performance* 1, 288-299
- [31] Einhorn, H.J. & Hogarth, R.: 1978. Confidence in judgement: Persistence of the illusion of validity. In: *Psychological Review* 85, 395-416
- [32] Kyburg, H.E. Jr. & Smokler, H.E.: 1964. *Studies in subjective probability*. New York: Wiley
- [33] Peterson, C.R. & Beach, L.R.: 1967. Man as an intuitive statistician. In: *Psychological Bulletin* 68, 29-46
- [34] Lichtenstein, S./Fischhoff, B. & Phillips, L.D.: 1977. Calibration of probabilities: The state of the art. In: H. Jungermann & G. de Zeeuw (eds.), *Decision making and change in human affairs*. Amsterdam: Reidel
- [35] Fischhoff, B.: Debiasing. In: D. Kahneman/P. Slovic & A. Tversky (eds.), *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge U.P., in press
- [36] Wyer, R.S.: 1974. *Cognitive organization and change: An information processing approach*. Potomac, MD: Lawrence Erlbaum Associates
- [37] Kahneman, D. & Tversky, A.: 1973. On the psychology of prediction. In: *Psychological Review* 80, 237-251
- [38] Slovic, P./Fischhoff, B. & Lichtenstein, S.: 1976. Cognitive processes and societal risk taking. In: J.S. Carroll & J.W. Payne (eds.), *Cognition and social behavior*. Potomac, MD: Lawrence Erlbaum Associates
- [39] Quinn, R.P. & Staines, G.L.: 1978. *The 1977 quality of employment survey*. Ann Arbor, MI: Institute for Social Research, University of Michigan

- [40] Murphy, A./Lichtenstein, S./Fischhoff, B. & Winkler, R.L.: 1980. Misinterpretations of precipitation probability forecasts. In: Bulletin of the American Meteorological Society 61, 695-701
- [41] Tversky, A. & Kahneman, D.: 1974. Judgement under uncertainty: Heuristics and biases. In: Science 185, 1124-1131
- [42] Miller, N.E.: 1944. Experimental studies of conflict. In: J.M.V. Hunt (ed.), Personality and the Behavioral Disorders, Vol. 1. New York: Ronald Press
- [43] Lichtenstein, S. & Slovic, P.: 1971. Reversals of preference between bids and choices in gambling decisions. In: Journal of Experimental Psychology 89, 46-55
- [44] Fischhoff, B./Slovic, P. & Lichtenstein, S.: 1980. Knowing what you want: Measuring labile values. In: T. Wallsten (ed.), Cognitive processes in choice and decision behavior. Hillsdale, NJ: Erlbaum
- [45] Slovic, P. & MacPhillamy, D.J.: 1974. Dimensional commensurability and cue utilization in comparative judgement. In: Organizational Behavior and Human Performance 11, 172-194
- [46] Grether, D.M. & Platt, C.R.: 1979. Economic theory of choice and the preference reversal phenomenon. In: American Economic Review 69, 623-638
- [47] Kunreuther, H./Ginsberg, R./Miller, L./Sagi, P./Slovic, P./Borkan, B. & Katz, N.: 1978. Disaster insurance protection: Public policy lessons. New York: Wiley-Interscience
- [48] Schwab, D.P./Olian-Gottlieb, J.D. & Heneman, H.G.: 1979. Between-subjects expectancy theory research: A statistical review of studies predicting effort and performance. In: Psychological Bulletin 86, 139-147
- [49] Staw, B.M.: 1977. Motivation from the bottom up. In: B.M. Staw (ed.), Psychological foundations of organizational behavior. Santa Monica, CA: Goodyear
- [50] Janis, I. & Mann, L.: 1977. Decision making. New York: Free Press
- [51] Corbin, R.M.: 1980. A theory of choice should not be based on choice alone. In: T. Wallsten (ed.), Cognitive processes in choice and decision behavior. Hillsdale, NJ: Erlbaum
- [52] Lakatos, I.: 1970. Falsification and the methodology of scientific research programmes. In: I. Lakatos & A. Musgrave (eds.), Criticism and the growth of knowledge. London: Cambridge U.P.
- [53] Payne, J.S./Braunstein, M.L. & Carroll, J.S.: 1978. Exploring pre-decisional behavior: An alternative approach to decision research. In: Organizational Behavior and Human Performance 22, 17-44

- [54] Beach, L.R. & Mitchell, T.R.: 1978. A contingency model for the selection of decision strategies. In: *Academy of Management Review* 3, 439-449

This Page Intentionally Left Blank

SOCIAL INTERACTION

This Page Intentionally Left Blank

CONDITIONS OF VIOLATING INDIVIDUAL RATIONALITY

WERNER H. TACK

*Universität des Saarlandes
Psychologisches Institut
F.R. Germany*

Introduction

This paper deals with a very special type of group decision making, namely decisions to settle conflicts of interest between at least three partners. The exclusion of the two-partner-case implies that

- (1) there is always room for non-trivial coalition formation, and
- (2) there are always reasons for- at least partial - cooperation due to the fact that a perfectly competitive situation for every pair at the same time is impossible.

Furthermore, the situations regarded here are *cooperative* in the sense that the partners involved have the possibility to communicate with each other and are allowed to enter binding agreements on how to solve the conflict at hand.

Formally, any of these situations can be represented by a set N of partners (persons or groups) - called "players" in game-theoretic terminology -, a series of sets $A_1, A_2, \dots, A_n$ of feasible options - one set for each player -, and a series of preference relations P_i - again one for each player - on the set of outcomes resulting from combinations

of chosen options. Introducing the additional assumption that all individual preferences follow the rule of negative transitivity and that - in case of an infinite outcome set - the outcome set contains a denumerable order-dense subset, individual preferences can be represented by payoff functions x_i such that for every outcome pair a, b

$$a P_i b \text{ iff } x_i(a) > x_i(b)$$

Thus, all situations regarded in this paper are representable by n -person-games in normal form.

In this context, the term "cooperative" simply refers to the fact that there is some kind of communication between the players involved, constituting a bargaining process before final decisions, and that binding agreements between some or all of these players on which options to choose are a possible result of such a bargaining process. As a matter of course, the possibility of binding agreements does not imply that any player *must* enter such an agreement. Thus, he can take the value of an outcome he can reach without actually cooperating with someone else as a yardstick for deciding which possible agreements are too low in value - according to his own preference system - to be acceptable. Such a behavior follows the simple principle not to accept an offer or to enter an agreement if there is a preferred outcome which can be reached by appropriate action without collaboration with any of the other players involved in the given conflict of interest. It is exactly this simple concept which is termed *individual rationality*.

Experimental Games

The experimental games used in our former research (KAUFMANN & TACK, 1975, LICHTENBERGER, 1975, TACK, 1980, 1981) can be taken as examples of the cooperative situations to be analyzed here. They are much simpler and more transparent than usual real life situations; we know a lot about their internal structure, and we have some experiences on the behavior of experimental subjects in these types of situations.

The first example is the "real estate brokers problem" which has been introduced as an experimental paradigm a couple of years ago (KAUFMANN & TACK, 1975). Imagine three real estate brokers - called Adams, Brown, and Clay. Each of them wants to sell a piece of land. The three estates are neighbouring; they all have pairwise common borders, and the whole situation looks like that depicted in figure 1. There are seven prospective buyers on the market. One of them wants to buy Adams' estate (A) and offers 7 units (maybe 700.000 German Marks or whatever one likes to use as the monetary equivalent of a unit). The second offers 2 units for Brown's estate (B), and the third 3 units for Clay's one (C). The next three prospective buyers are interested in pairs of estates only. One offers 10 units for A+B, one 7 units for B+C, and one 12 units for A+C. Finally, there is a company who wants to buy the whole area; they are willing to pay 14 units for all three pieces of land together.

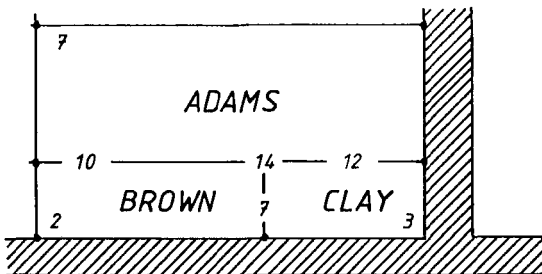


Figure 1: Representation of a sample situation from experiments using the "real estate brokers problem".

These values constitute a function from the set of possible *coalitions* S of brokers into the reals, if - as usual in this context - the term "coalition" is used for any nonempty subset of players, including - for the sake of simplicity - isolated players as "one person coalitions" and - at the other extreme - the "grand coalition" comprising all partners. Such a function is called a *characteristic function* and it is usually taken to be *superadditive* which means that for every pair of disjoint coalitions the function value of their union is at least as great as the sum of the values of the original coalitions; formally:

for every $S, T \subseteq N$ with $S \cap T = \emptyset$:

$$v(S \cup T) \geq v(S) + v(T).$$

In the brokers situation, superadditivity is a necessary requirement to make things plausible. If there is - for instance - an offer of 10 units for the areas A and B together, and another of 3 units for C alone, somebody interested in the whole area must offer at least the sum of these values - namely 13 - in order to be seriously taken into account as a possible client.

The brokers are the players engaged in a conflict of interest. They have to solve the problem, to whom to sell and how to share the resulting gain. Introducing students as experimental subjects playing the roles of the brokers gives an example of one of the experimental situations used to collect data on bargaining behavior, coalition formation, and conflict resolution. Experiments of this type are rather usual; except that other researchers mostly give to their subjects the abstract values of a characteristic function without such a nice cover story, and with a second deviation from the paradigm introduced here which will be mentioned later in another context.

For this type of situation, the concept of *individual rationality* is extremely simple:

Nobody should agree on a solution which gives him less than the amount of units he can gain by selling his real estate alone to

the one person who is interested in this and only this area.

Things are a bit more complicated in another class of experimental situations, the so-called "three-person-matrix-game". Three subjects are seated at a round table. Each of them faces a display showing three tilted 3x3-matrices like those depicted in figure 2. Each of the 27 cells of these matrices represents a possible outcome of the situation which is symbolized by three numbers. These numbers specify amounts of money. An outcome is realized by paying the amount given by the top number to the subject facing this display, while the amounts given by the bottom-left and bottom-right numbers are paid to the players sitting on his left- and on his right-hand-side respectively.

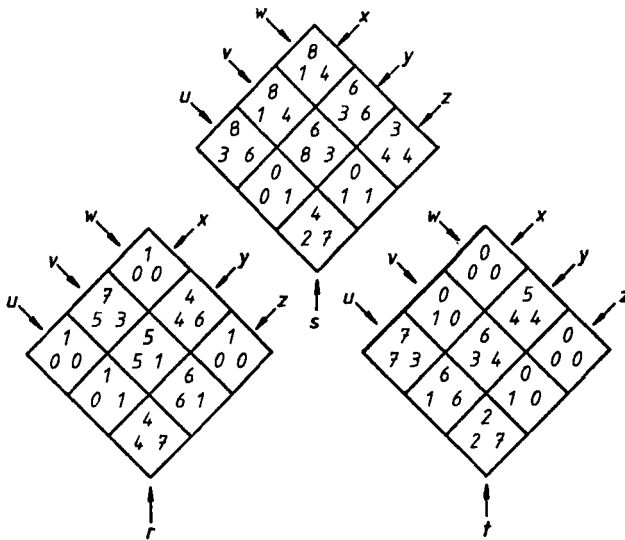


Figure 2: Example of the representation of an outcome configuration as displayed to an experimental subject in a matrix-game situation.

The feasible options of each player are the three matrices on his display (marked "r", "s", and "t" on figure 2). He can choose one of them, thereby restricting the set of possible outcomes to those given by the chosen matrix. Information on the display faced by the other subjects is arranged such that an option of the partner sitting on the left-hand side corresponds to one of the rows from upper-left to lower-right (marked "u", "v", and "w"), while the right-hand player can restrict possible outcomes to those in one of the rows going from upper-right to lower-left (marked "x", "y", and "z").

Imagine - for example - that the player facing the layout given in figure 2 and his right-hand partner agree to choose the middle matrix (s) and - on the partner's display - a matrix corresponding to the upper row from upper-right to lower-left (x). Then, only the three outcomes in the upper row from upper-right to lower-left in the middle matrix (the s-x combination) are still possible. If the third player decides to choose on his display a matrix which is represented here as the lower row from upper-left to lower-right (u), s-u-x marks the position of the final outcome. The person facing the layout depicted in figure 2 has earned 8 units while the other two players gain 3 and 6 units.

Again, an outcome in this type of situations is called *individually rational* if nobody gains less than the amount he can assure for himself by choosing appropriate actions without collaboration with any other partner. But how can a person - confronted with a situation of this type - enforce something? There are at least two answers to this question, corresponding to the distinction between alpha- and beta-rationality introduced by AUMANN and PELEG (1960) more than 20 years ago. This distinction can be embedded into a more general concept of situational "control"; this approach will be used here to clarify the distinction between different kinds of individual rationality.

Control Systems

It is a central aspect of all the social situations we are regarding in this context, that every partner engaged in such a situation - and every coalition too - has possibilities to restrict by appropriate actions the

set of feasible outcomes; he has - at least partial - control. The system of outcome subsets generated by the different possibilities of restriction which are available to a partner is called a *control system* C_i of the partner $i \in N$. Thus

$$C_i \subseteq P(A),$$

where $P(A)$ denotes the power set of the unrestricted outcome set A . It is necessary to introduce some additional properties of control systems.

- (1) It should *always* be possible to refrain from any action which makes some outcomes unreachable; thus - there is always a possibility to force no restriction at all. This means:

$$A \in C_i \text{ for every } i \in N.$$

- (2) It should *never* be possible to make all outcomes unreachable, or - in other words - to force the situation into an eternal state where it never can end. Formally:

$$\emptyset \notin C_i \text{ for every } i \in N.$$

- (3) It should *always* be possible to restrict the set of available outcome-restricting actions, which means that

$$\begin{aligned} &\text{if } E, F \in C_i, \\ &\text{then } E \cup F \in C_i \text{ for every } i \in N. \end{aligned}$$

In fact, control systems should be introduced by stating some more properties than these three, but the others have to do with systems of coalitions consisting of more than one partner only and are - therefore - of no special interest in the relevant context of the considerations following in this paper.

Given a control system C_i of an individual player i , the best value i can enforce equals the maximum of the minima of outcome-values from the several

outcome sets belonging to the system. By restricting the set of feasible outcomes to a specific subset B belonging to C_i , he can guarantee to reach the minimum value over all outcomes in B . Choosing for restriction that subset B' which gives the best guarantee in the sense that there is no other subset with a higher minimum value, the maximum of the minima becomes a lower bound for the values of all the outcomes which are still feasible. Formally, i can enforce

$$\text{Max} \{ \text{Min} \{ x_i(b) \mid b \in B \} \mid B \in C_i \}.$$

Thus, outcome b is *individually rational* if and only if for every $i \in N$:

$$x_i(b) \geq \text{Max} \{ \text{Min} \{ x_i(a) \mid a \in B \} \mid B \in C_i \}.$$

Going back to the matrix situation, one can introduce at least two different control systems. A very simple type of control consists of restricting the own set of options. If a player facing the layout given in figure 2 decides not to take the right matrix, he restricts the set of feasible outcomes to those given by the two other matrices. Thus, the set of these outcomes is an element of a control system called *alpha-control system*. By deciding to take the lower left matrix, he can enforce to gain at least 1 unit (in the other matrices, he can - possibly - get nothing). Thus, 1 unit is the value he can enforce, and - in order to be individually alpha-rational for him - an outcome must give him at least 1 unit.

Beta-control is another story. In his case, the person facing the layout given in figure 2 decides on which matrix to choose dependent on the choices of his partners. He reacts to the decisions of the others by choosing a function from the set of combinations of their options into his own set of available alternatives. One such function determines the set of feasible options depicted in figure 3, which guarantees an outcome with at least 3 units for the person facing this layout. Thus, under *beta-control* 3 units is a value he can enforce, and - in order to be individually beta-rational for him - an outcome must give him at least 3 units.

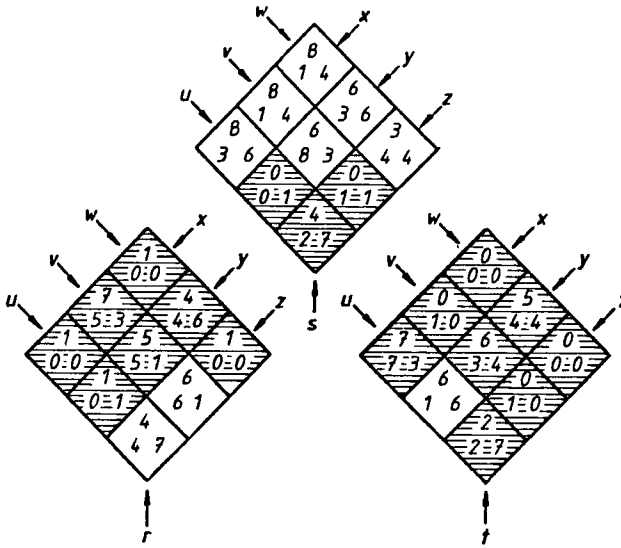


Figure 3: Outcome configuration of a matrix-game showing a set of outcomes (unhatched) belonging to the beta-control system of the player facing this layout.

The problem with beta-rationality is that the realization of beta-control presupposes a fixed time order of final decisions and full information about the decisions of the other players. In the experimental games run at our laboratory (LICHTENBERGER, 1975; TACK, 1980, 1981), there was no fixed time order, but during bargaining we heard very often sentences like

"If you do this, I'll do that, otherwise..."

If people accept such announcements as serious ones, they behave as if all of them could simultaneously exert beta-control, which is factually impossible. AUMANN and PELEG (1960) suspect that while alpha-control is obviously more plausible it may turn out that beta-control is much more important for real cooperative behavior. We have some empirical evidence

supporting this conjecture.

Results

Looking at experiments using characteristic function situations - like the "real estate brokers problem", but usually without such a nice cover story - the relevant literature reveals a rather strange situation. It seems to be usually taken for granted that subjects in experimental games agree on individually rational outcomes - maybe with a very small number of exceptions -, but nearly all experiments (e.g. RIKER, 1967; LAING & MORRISON, 1973; RAPOPORT & KAHAN, 1976; ALBERS, 1978; RAPOPORT & KAHAN, 1979) have a 0-percent violation of individual rationality due to the fact that the experimenters don't give their subjects any chance to violate this principle. The characteristic functions have zero-values for the 'one-person coalitions', which simply means that there is no possible outcome over which an individually enforceable outcome is preferred.

Looking at experiments using the real-estate-brokers paradigm, there are - up to now - data from 228 single experiments using 25 different characteristic functions. Most of these data are published (KAUFMANN & TACK, 1975; TACK, 1980), some of them have been collected by students as part of their seminar work (reference note). The proportion of experimental bargaining situations with final agreements on individually non-rational outcomes is between 2 and 3 percent only. Thus, violations of individual rationality seem to be a very rare event. This can lead to the conclusion that such a spurious phenomenon does not deserve any special interest, that it can be put aside as one more manifestation of simple faults and errors which occur in every context where fallible human subjects are involved. But, when looking at the distribution of the violations of individual rationality over the different characteristic functions, things are starting to become interesting. All violations are concentrated at about 10 percent of all characteristic functions. Taking alone these functions into account, the proportion of violations goes up to 19 percent. First consequence:

Whether violations of individual rationality are a rare event or not

depends on the choice of characteristic functions used for experimentation.

Asking for the special properties of characteristic functions which seemingly attract violations of individual rationality shows that they are marked by large differences between the individually enforceable values, combined with a small difference between the total amount available for distribution in the grand coalition and the sum of the individually enforceable values. In terms of our real-estate-brokers paradigm: The probability of violations of individual rationality goes up if there are marked differences in the prices offered for single pieces of land and a very small additional gain which can be earned by selling the whole area to the one prospective buyer who wants to have all three pieces together.

In such a situation, there is a very small monetary incentive for the formation of a grand coalition; the grand coalition is rather worthless under payoff maximizing considerations. Thus, if - nevertheless - the grand coalition is formed, there must be another reason to do this. One possible explanation is that - in this special type of situation - subjects use the grand coalition to diminish interindividual differences in enforceable results. It is a well known fact, that equity considerations play an important role when people have to share a certain amount of money, and if there is no power or enforceability enabling strategic considerations of the kind regarded in this paper. If coalitions are rather useless under strategic aspects, they probably become useful for other purposes, e.g. for "greater equality" in the sense of diminishing differences. On the basis of this interpretation, the connection between enforceable values is a kind of secondary effect. The greater the variance of individually enforceable results, the higher the probability that tendencies in direction to an equal share automatically violate individual rationality. For better information on this phenomenon, it is necessary to design an experiment controlling as an experimental variable the number of players for whom an equal share of the grand-coalition value is individually rational.

Data on violations of individual rationality in matrix game situations

are - for many reasons - much more informative than those from brokers problem sessions. Mainly in connection with a project run by LICHTENBERGER (1975) data have been collected from 390 bargaining sessions using 32 different outcome configurations. This data file contains 23 violations of individual rationality. Before calculating any relative frequencies, one has to subtract 85 experiments with payoff configurations in which every outcome is individually both alpha- and beta-rational. What remains are data from 305 experiments showing 22 violations of alpha-rationality (that is 7 percent) and 23 violations of beta-rationality (that is about the same percentage).

Since every matrix situation has a finite number of outcomes, one can count the different numbers of available outcomes meeting certain criteria. In the whole set of outcome configurations - excluding those in which every outcome is alpha- as well as beta-rational - 41 percent of all the available outcomes are individually alpha-rational while 33 percent are beta-rational. The extremely small difference between the frequencies of alpha- and beta-rationality among the *chosen* outcomes is not deducible from a similarly small difference in the set of *available* ones. Beta-rationality implies alpha-rationality; nearly all - with one exception only - chosen individually rational outcomes comply with the stricter laws of beta-rationality. This is one of the findings supporting the conjecture that, despite the greater plausibility of alpha-control, beta-control has a higher explicative value for real bargaining behavior and its results. In our data files, there are some outcome configurations in which individual alpha- and beta-rationality cannot be differentiated - each alpha-rational outcome is beta-rational too - and some other configurations in which *every* available outcome is individually alpha-rational, such that only the *additional* requests of the beta version specify beta-rationality. In both subgroups of experiments, the percentage of violations of individual beta-rationality is the same. This is another evidence for the fact, that it is the concept of beta-rationality which really counts if we are dealing with empirical data. One of the reasons for not recognizing this earlier could be that nearly all experimentation has used side payment situations in which alpha- and beta-control cannot be distinguished from each other.

If one takes a closer look at the violations of individual rationality neglecting the alpha-beta-distinction which is relevant for one single case only, there is a clear and eye-catching difference between two kinds of violations. The first kind is marked by the fact that the player for whom the outcome is not individually rational is always isolated at the end of the bargaining process; he is never a member of a larger coalition. Further, violations of this type are scattered about the different outcome configurations, showing no recognizable correlation with any property of the configurations. This indicates that there is no consensus to sacrifice individual rationality in order to reach some other common goal; it is a single person's decision which leads to a violation of individual rationality. It is probably not too complicated to construct a number of heuristic decision rules leading to exactly this kind of phenomenon. One could start by stating assumptions on a person's assumptions concerning his partners decision behavior or - as another possibility - by introducing an appropriate mechanism which overlooks certain combinations of the other persons' options or certain available outcomes. Further pursuit of this type of interpretation presupposes a greater number of realized violations of this kind; thus, it is probably advisable to postpone such an approach until there is some more information on this kind of violations.

The second type of violations looks totally different. Here, it is an agreement of all persons engaged in the situation which leads to the violation of individual rationality; it is never a single person's but always a common decision to choose an outcome which is not individually rational. Violations of this kind are no longer unsystematically scattered over the whole set of outcome configurations; their distribution is concentrated at some very few configurations only. Non-rational outcomes of this type always show differences between the amounts paid to the partners which are definitely smaller than the differences between the amounts enforcable by individual control. The configurations with a high rate of agreements on individually non-rational outcomes of this type show distinct differences between the individually enforcable values combined with small or even no advantages of entering some coalition.

The properties of configurations which seemingly attract this type of individually non-rational outcomes can be illustrated by a configuration which is rather extreme in all these respects. The person facing the layout given in figure 4 can decide to take the left matrix and enforce to

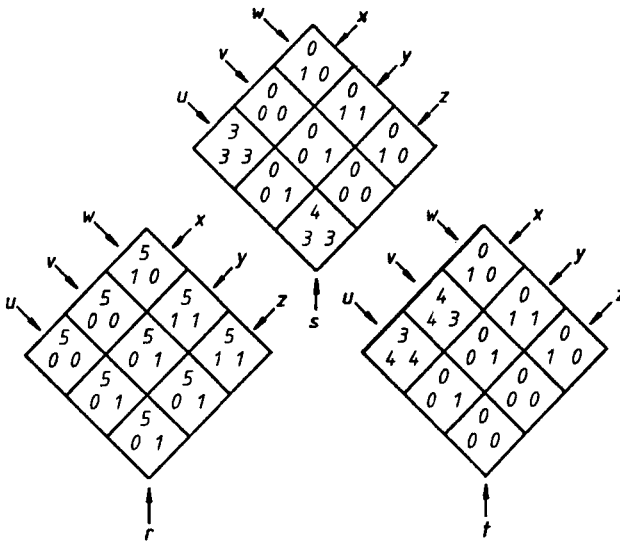


Figure 4: Special case of an outcome configuration leading to an exceptionally high proportion of violations of individual rationality.

reach 5 units by exerting alpha-control. The maximum enforcable by each of his partners is 1 unit only. Beta-control does not create any advantage to anybody. Since 5 units is the maximum for the player facing this layout, there is no incentive to enter a coalition. Each of the others can gain from a grand coalition only, but has to take into account that the partner facing the layout from figure 4 has no reason to enter

such an agreement. Under gain maximizing considerations, one expects that the bargaining ends with three isolated partners, each of them getting what he can enforce and - thus - contributing to an individually rational outcome. In 10 out of 15 experiments we run using this configuration, something totally different happened. The partners entered a grand coalition and agreed on an outcome being not individually rational but diminishing the differences between enforceable values. In fact the grand coalition's choice was to give 4 units to two of them and 3 units to the third or - but in one case only - to realize the non-Pareto-optimal equal share solution which gives 3 units to each of them.

Comparing these results with those from the brokers problem experiments shows that the second kind of violations encountered when running matrix experiments is again a grand coalition's decision in situations with marked interindividual differences which serves the purpose to diminish these differences and to create a solution closer to an equal share. The main results of the analysis of violations of individual rationality are:

- (1) There is evidence that the stricter properties of beta-rationality are more useful in explaining behavior in experimental games than those stemming from alpha-control considerations only. This is not observable inside-payment situations which do not allow to differentiate between alpha- and beta-control.
- (2) There are two clearly differentiable types of violations of individual rationality. One type consists of violations due to - in the sense of rationality considerations - un-adequate decisions of a single and isolated partner. These violations show no recognizable dependencies on properties of the given outcome configuration; they are not observable in side payment situations in which an isolated partner automatically gains his individually enforceable value.
- (3) The second type of violations - observable in side payment situations as well as in experimental games without side payments - is based on grand coalition decisions. It is prominent in such situations where the individually enforceable values are very different and where the

strategic incentive of entering a coalition is rather small. These violations always diminish the interindividual differences leading to outcomes which are not individually rational for those partners being strong according to what they can enforce without actually cooperation.

Looking at other kinds of "wrong" or "false" human decisions, it is probably informative that our research shows the fruitfulness of considering systematic differences between several 'types' of errors which not only show different dependencies on situational variables but serve different purposes too. Looking at future research on violations of individual rationality, it is probably one of the most important results of our analysis that we now know how to construct situations leading to a higher number of violations. This allows the design of experiments being more informative with respect to these violations and to leave the field of rather generous hypotheses formation based on extremely small numbers of cases observed in experiments designed for other purposes.

Reference note

Anstadt, T./Niessner, M./Schaefer, H.J. & Thraem, S.: 1980. Entscheidungsverhalten bei 3-Personen-Spielen mit Seitenzahlungen. Fachrichtung Psychologie der Universität des Saarlandes, Saarbrücken

References

- Albers, W.: 1978. Bloc forming tendencies as characteristics of the bargaining behavior in different versions of the Apex game. In: E.H. Sauermann (ed.), Coalition forming behavior. Tübingen: Mohr
- Aumann, R.J. & Peleg, B.: 1960. Von Neumann-Morgenstern solutions to cooperative games without side payments. In: Bulletin of the American Mathematical Society 66, 173-179
- Kaufmann, M.A. & Tack, W.H.: 1975. Koalitionsbildung und Gewinnaufteilung bei strategisch äquivalenten 3-Personen-Spielen. In: Zeitschrift für Sozialpsychologie 6, 227-245
- Laing, J.D. & Morrison, R.J.: 1973. Coalitions and payoffs in three-person sequential games. In: Journal of Mathematical Sociology 3, 3-25
- Lichtenberger, J.: 1975. Drei-Personen-Spiele mit Seitenzahlungen. Doctoral dissertation, Universität des Saarlandes, Saarbrücken

- Rapoport, A. & Kahan, J.P.: 1976. When three is not always two against one: Coalitions in experimental three-person cooperative games. In: *Journal of Experimental Social Psychology* 12, 253-273
- Rapoport, A. & Kahan, J.P.: 1979. Coalitional power and payoff disbursement in three-person cooperative games. Chapel Hill, N.C.: L.L. Thurstone Laboratory, University of North Carolina, Report No. 159
- Riker, W.H.: 1967. Bargaining in a three-person game. In: *The American Political Science Review* 61, 642-656
- Tack, W.H.: 1980. Cooperative conflict resolution. A formal approach and some empirical results. In: E.D. Lantermann & H. Feger (eds.), *Similarity and choice: Papers in honour of Clyde Coombs*. Bern: Huber, 290-326
- Tack, W.H.: 1981. Domination in experiments on cooperative games without side payments. In: G.E. Lasker (ed.), *Applied systems and cybernetics*, Vol. 2. New York: Pergamon Press, 728-732

This Page Intentionally Left Blank

BIASES IN GROUP DECISION MAKING

HELMUT W. CROTT

JOHANNES A. ZUBER

*Universität Freiburg
Psychologisches Institut
F.R. Germany*

1.1 Introduction

The following study intends to explain *Polarization* in group decision making by group decision rules, which take into account the rankorder of individual preferences over alternatives.

The research on *Group Choice Shift* or *Group Polarization* (alias risky shift/cautious shift) was initiated by STONER (1961), who compared the risk taking of individuals and groups. In his experimental setting he asked the participants first to respond individually to a series of story problems. Each problem described a risk involving dilemma situation. In a second phase participants assembled as groups and discussed the problem until they came to an agreement. STONER's striking finding was that the group decisions were usually more risky than the average individual decision. In addition, it was found that a brief period of discussion, followed by an individual decision-making will also cause a shift in the group average.

Although the average shift in probability was only about .10 or less, this finding, which was unexpected according to the existing theories, has raised interest among social and organizational psychologists. Besides its role in the understanding of human behavior in interaction situations, the 'risky shift' appears of importance for many organizational situations in which decisions are made by teams. It is easy to imagine the severe results that may ensue when business committees, military decision groups, counselling

groups and juries first hold a group discussion and then make a joint decision.

Therefore, it is not surprising that STONER's initial study was followed by a wave of investigations of group risk-taking. The theoretical explanations for that phenomenon based on experimental research can only be mentioned briefly in the given text (for more detailed information see summaries by LAMM & MYERS 1978, and MYERS 1982).

The following factors have been taken into account:

1. In groups, the *responsibility* for a decision is *distributed* among the participants, thus the decision may tend to be more risky.
2. At the outset *Ss are not familiar with the material* and thus tend to be very cautious. During the discussion *Ss* exchange information about the matter and in so doing are able to perceive the situation more realistically, which usually leads to a higher risk taking.
3. The *opinion leaders* in the group *are usually more risky*.
4. The *group atmosphere* is comparably more positive and thus *creates optimism*.
5. *Risk is a value* in our society, therefore, the majority will already lean towards being more risky and social comparison will cause the cautious minority to join the majority.

Some evidence has been provided for each of these hypotheses. However, most of these explanations can be found faulty, since it has been observed that some material creates more conservative than risky shifts (as was already found with items in STONER's study). On the other hand, at least statements 3. and 5. can be adapted in order to account for conservative shifts as well. This can be seen in that the leader usually makes extreme decisions, either in the direction of being conservative or risky and in the fact that cautiousness, as well as riskiness, is a highly-esteemed value in our society. It was also found that group shifts appeared with regard to non-risk situations. Therefore, in the late sixties the term risky shift was substituted by 'group-(response)-shift' and 'polarization' (MOSCOVICI & ZAVALLONI 1969). Together with the introduction of these more general terms,

the theoretical explanations had to be reformulated in order to cover shifts in attitudinal, value-related and judgmental items.

1.2 Definition of Shift and Polarization

MYERS (1982, p. 127) describes group polarization as the fact that "the average post-discussion response (*and group decision*) will tend to be more extreme in the same direction as the average of the pregroup responses". The question here is how the *direction* is determined with regard to the midpoint of the scale or with regard to the average individual response of all Ss taking part in the experiment? In experimental research both criteria are used to determine the direction of shift. These two criteria for polarization have, however, certain shortcomings. The midpoint of the scale does not always represent the psychological midpoint and it often does not even exist (e.g., in estimation problems which have a natural zero point but not always a natural maximum). In regards to the second criteria the average first individual preference is dependent on the population and thus, it is affected by sampling biases. As a third method, especially suitable to test the degree of polarization not caused by the Ss' tendency to decide according to a majority rule, MYERS (1982, p. 139) suggests to compare the median of the individual responses with the group decision or the post discussion responses. This method corresponds better to the original definition of shift. The arithmetic mean within the group is taken as criterion and the location of the median as an indicator of the skewness of the distribution.

1.3 Social Decision Rules

In the discussion above, it was already assumed that the implicit rule which groups apply in order to reach a decision might cause the difference between the arithmetic mean and the group decision. The importance of decision rules (amalgamation rules, social decision schemes) was first pointed out by economists in discussing problems of social welfare (see e.g. ARROW 1951, BLACK 1948). Social psychologists became aware of the problem through the research on Social Decision Schemes (see DAVIS 1973). Theories on social decision making and relevant research will not be detailed in this paper. Within the given framework, some features of social decision rules which are important

for the following analysis of polarization and shift will be discussed. Decision rules usually predict the outcome of a group decision on the basis of the individual preferences of the participants by integrating the individual preferences into a collective decision according to the given rule. Most decision rules investigated empirically thus far, take into account only the first preferences of the individuals. Some decision rules predict a distribution of group decisions. Others make exact point predictions or specify the range in which the group decision is most likely to fall.

Examples for distributional group decision rules are as follows:

- *Proportionality rule* - The distribution of group decisions is proportional to the distribution of the first preferences of the individuals.
- *Equiprobability rule* - Any alternative preferred by at least one individual will have the same probability of becoming the group decision.

The following three are examples for point/range predictions:

- *Arithmetic Mean* - Average of all first preferences of the individuals.
- *Majority Decision* - The absolute majority for one of the alternatives.
- *Plurality rule* - An alternative will result from a group decision which has the relative majority, i.e. the modal alternative.

Naturally, it depends on the task if a model makes sense as a group decision predictor. Whereas the Proportionality rule and the Equiprobability rule might apply for any problem, rules like the Arithmetic Mean and others make sense only if the object of decision is quantifiable on the level of an interval scale. Some decision rules do not lead to unique decisions. In these cases, either the range of outcomes is predicted or a *secondary decision rule* is introduced. Consequently, sometimes in small groups there is more than one Plurality (i.e., two or more modal values in the distribution of first preferences). When this occurs, one might predict a set of outcomes (all alternatives with a plurality are equally likely) or one might predict with another secondary decision rule that the outcome will be the median plurality or will be located in the middle between opposing pluralities.

1.4 The relation between Social Decision Rules and Polarization

Research on shift and polarization has implicitly taken the arithmetic mean of the individual responses as the natural candidate for the group decision. The fact that the actual group decision deviated from the arithmetic mean was therefore interpreted by psychological effects like the effects of group atmosphere, leader characteristics, information exchange effects and societal values (as described in the introductory section). However, it is easy to verify that most decision rules will lead to a shift in relation to the arithmetic mean, when the distribution of the original individual preferences within the group is not completely symmetric. Moreover taking small groups within a population will usually lead to asymmetric distributions, even if the population's distribution is symmetric. So polarization is likely to occur as a consequence of the decision rule. Taking that into account the central question for the following empirical investigations can be formulated: 'Do social decision rules account for all the polarization' or alternatively 'Is there substantial polarization left after the application of a decision rule?'.

1.5 Some remarks regarding the concept of the following empirical studies

The empirical studies reported in the following section were originally designed to analyse which decision schemes fit best on collective decision. However, for the reasons discussed above, the studies will suit a reanalysis in terms of the polarization effects.

Ad hoc groups have usually been taken as the basis of analysis for research on decision schemes. However, the distribution of individual preferences in ad hoc groups does not very often allow for differentiation of different social decision rules. It may happen with ad hoc groups that according to the overall statistics all decision schemes predict equally well, although in any single group the decision can be best explained by a certain point-prediction model. Because of the difficulty to differentiate between certain models by using ad hoc groups it is more powerful to construct groups in a way that the individual preferences fulfill certain distributional characteristics. This has been done in study II and III. That would not have been

necessary for the analysis of the polarization effect, but this does not, however, affect the validity of that analysis. In all three studies hypotheses about the best fitting social decision scheme were formulated in advance, i.e., depending on the task, variations of the majority decision rule were assumed to be best fitting. The basic idea behind the majority decision rules applied in the three studies is the assumption that group decisions will fall on the first preference of the median individual in tasks with underlying scales of (at least) ordinal character. As BLACK (1948) has proved, that is in fact the case when the preference curves of the Ss over the scale points are single peaked, and when the decision rule is the pairwise comparison of alternatives by majority voting. Single peaked means that the individual has a preference point (bliss point, ideal point) on the scale, and that his degree of satisfaction with other scale points decreases continuously left and right of that preference point. As will be demonstrated later this assumption is not trivial, especially not for risk-taking situations. If the single peakedness condition does not hold then, a majority voting over the alternative decision might lead to the group decision different from the median. This differs from most other decision schemes in that those aggregation rules have to take into account not only the first preferences but the complete preference rankorders. The decision rules predicted as best fitting are given in a brief description of the experiments in the next sections. In order to analyse the polarization effect in terms of social decision schemes, we have to compare actual polarization and the polarization predicted by the social decision rule. As a basis we take the decision schemes that proved as best fitting, assuming that those rules approximate best what Ss really used as a collective decision device. The decision rules assumed as best fitting in advance indeed showed the smallest deviation from the actual group decision. In addition we analysed various other prediction schemes, like geometric mean, arithmetic mean, plurality, proportionality and equiprobability (CROTT 1965, CROTT/SCHERMER/KNOLL & THOENE, in prep., and CROTT/SCHERMER & ZUBER, in prep.).

1.6 General Hypotheses for Experiments I to III and technical remarks regarding the analysis and presentation of data

Our general idea says that in cases where the decision alternatives can be

rank ordered at least on an ordinal scale the collective decision is reached by some kind of majority voting. If there is an absolute majority for one alternative (more than 50 % of group members give the first rank to the same alternative) then this alternative will result as a group decision. In the more critical cases, mainly investigated in the following experiments in which no absolute majority exists for one single alternative, the result of the majority voting depends on the structure of the individual preference orders:

1. If all individual preference orders are single peaked, the result of the majority voting will be the alternative most preferred by the median individual. If there are two median individuals a secondary decision rule has to be applied (e.g., midpoint of the range of the two median individuals).
2. In cases in which the single peakedness condition is not fulfilled for at least one individual in the group a *paired comparison by majority voting over all pairs of alternatives* might lead to a result differing from the median. Hence, when single peakedness of individual preference curves can't be assumed to hold generally, we have to compute the pair-wise majority voting result for each groups of Ss separately. - In some cases this procedure would result in intransitivities in the group's rankorder of the alternatives. Then the group could not be able to decide for one of the alternatives as a group decision. Again for those cases we have to establish a secondary decision rule.

Experiment I is a reanalysis of a study by CROTT (1964/65) with a judgmental task (estimation of the number of points on a slide after a short exponation). In Experiment II and III (CROTT et al., in prep. a,b) Ss have to decide on bets with equal expected value, varying in the degree of risk. It is the aim of all three analyses to demonstrate that there exists shift and polarization when conventional measure, namely the *deviation of the arithmetic mean*, is taken. At the same time we claim that shift and polarization reduce considerably when an appropriate decision rule is applied. In order to save space we only give a brief description of the experimental method. More detailed information is (or will be) given in the original articles. For the same

reason we mainly present the more general polarization data. The main results regarding left/right shift or risky/conservative shift are mentioned in footnotes under the polarization tables. The model testing approach applied here, implies maintaining the model, when deviations from the prediction are insignificant. Following a convention, the 20 % level of significance will be chosen when significant results are not expected according to our hypotheses. Otherwise the 5 % level is applied. This approach is conservative in making it difficult to get supportive evidence in favor of the hypotheses.

2. *Experiment I: Group Decision under different motivational conditions*
(CROTT 1964, 1965)

2.1 *Brief description of the experimental conditions*

Subjects and conditions of participants: Volunteering 72 male and 72 female students in 36 groups of four Ss. All members of a group were of same sex.

Tasks:

- a) Individual estimation of randomly distributed points on a slide after a short exposition (10 sec.).
- b) Giving a group judgment on the number of points.

Procedure:

Ss denoted their individual estimations, then the experimenter announced them in a random order. Then Ss wrote down new estimations without seeing the slide again. This was repeated until all estimations agreed within a range of 4 points.

Motivational conditions:

Motivation A: We want to know if groups reach better estimations of the number of points than individuals do on the average. Therefore please be cooperative to come to a good group result.

Motivation B: This experiment has to do with the relationship between intelligence and the ability to judge facts correctly. We want to know if individuals with the better estimations influence the group decision process more than individuals with worse estimations.

Under both conditions the experimenter announced that he would not be able to tell them the real number of points in order to prevent that Ss would talk about it to later coming Ss.

Experimental factors:

Sex and motivational conditions.

Social Decision Schemes applied:

Primary SDS: Median

Secondary SDS: Cooperative Motivation: Midpoint of the median range
(Median/Midpoint)

Competitive Motivation: Arithmetic mean if within the median range or that bound of the median range which is closer to the arithmetic mean
(Median/Min. Var. Distance).

Relevant Hypotheses:

In a paired comparison over the original individual estimations the majority rule could lead to a group decision within the *range of the two median Ss*, regardless of the motivational conditions. The secondary decision scheme differs depending on the motivational condition (as described above). We expect that no polarization will exist after application of the appropriate decision scheme.

2.2 Results and discussion for Experiment I

Since the Ss' estimations are given in numbers here, the single peakedness of the individual preference orders appears to be a natural property of the task. With four Ss in a group there exist two median points. Therefore a secondary decision rule has to be defined. Under the cooperative motivation we postulate that Ss will soon agree on the midpoint of the median range as an estimate. The secondary decision rule for the competitive motivation follows a hypothesis by HOFFMAN, FESTINGER and LAWRENCE (1954). It states that because of social comparison processes Ss will tend to minimize variance between the participants. Therefore we assume that Ss will agree on the arithmetic mean when it lies within the median range or, otherwise, on that bound of the median range that is closer to the arithmetic mean.¹⁾

The social decision rule might therefore state 1. the median range 2. within the median range, the midpoint of the two extreme Ss or that bound of the median range which is closer to it. We will however not present the data for that analysis here, since differences to the decision scheme suggested by social comparison theory are negligible.

The results in table I/1 and I/2 show that in the given judgmental context

1) A plausible alternative secondary group decision rule would lead to almost identical results, namely that the two extreme Ss bargain much tougher for the group result than the middle Ss which are already close to their original estimation by the primary decision rule.

Table I/1: Frequency analysis of *Polarization* (from the Arithmetic Mean in direction to midpoint of the Median range) +)

	Polarization		Polarization	
	+	-	+	-
	Arithmetic Mean		Median/Midpoint	
Cooperative Motivation	15 (13.5)	3 (4.5)	11	7
			Median/Min.Var.Dist.	
Competitive Motivation	12 (13.5)	6 (4.5)	7	11
	27	9	18	18

χ^2 (Polarization Arithm. Mean 27+ vs. 9-) = 9.0, df = 1, $p < .01$
Left/right shift is not significant.

+) No differences in frequencies between male and female groups.

Table I/2: Deviation analysis of Polarization (from the Arithmetic Mean into the direction of the midpoint of the median range) +) ++)

		Arithmetic Mean	Appropriate SDS: Median/Midpoint
Cooperative Motivation	$\bar{X}$	79.18	8.82
	s	147.28	30.85
	t	2.28 ^{.05}	1.21 ^{.30}
			Median/Min. Var. Distance
Competitive Motivation	$\bar{X}$	34.63	-5.04
	s	57.62	46.25
	t	2.55 ^{.05}	-.46 ^{.70}
Sums over both conditions	$\bar{X}$	56.91	1.89
	s	111.83	39.31
	t	3.05 ^{.01}	.29 ^{.70}

+) No differences between means for male and female Ss.

++) Regarding left and right shifts there is left shift from the Arithmetic Mean for motivation A and B together, with $t = 1.62$, $df = 35$, $p < .20$ and almost no shift for the appropriate SDSs.

there exist strong polarization effects according to the arithmetic mean. In that context left/right polarization from the arithmetic mean can hardly be accounted for by the usual psychological explanations provided for risk and attitudinal changes since it is just a change in the estimation of the numbers of points. The polarization is seen in the frequency table I/1 as well as in the analysis of deviations in table I/2. Polarization effects can, however, be reduced to almost zero when the hypothesized decision rules are applied.

3. *Experiment II: Social Decision Scheme and Group Polarization in Group Decisions among bets (CROTT, SCHERMER, KNOLL & THOENE, in prep.)*

3.1 *Brief description of the experimental conditions*

Subjects and conditions of participation:

260 male students of economy and law. Each S got 5 DM for participation. Ss were informed, that they could possibly win additionally a considerable amount of money. For each session between 15 and 25 Ss were invited.

Instructions, Tasks, and Procedure:

- a) Explanation of probabilities in terms of relative frequencies and the expected value as a long run expectation. Explanation of the meaning of individual preferences over bets with equal expected values. It was already mentioned to the Ss here, that they afterwards had to discuss their preferences with others to come to a group decision.
- b) Measurement of individual preferences for the following four bets:

	(1)	(2)	(3)	(4)
Probability	.10	.30	.50	.70
Payoff I	66.-	22.-	13.20	9.43
Payoff II	21.-	7.-	4.20	3.-

The Payoff conditions served as two different experimental conditions.

- c) Arrangement of Ss into groups which fit into one of the following distributions of occupancy numbers:

Images				Mirror Images			
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	1	2	0	0	2	1	1
1	2	2	0	0	2	2	1

- d) Group discussion and decision: Ss had to come to a group decision among the bets (after at most 15 minutes of discussion).
- e) Second measurement of individual preferences.
- f) Playing out of the bets: The chosen bet was played out once for the whole group and the amount of money of the bet or nothing was paid to the Ss depending on the result.

Experimental Factors:

The two payoff conditions under section Instructions b) and the four distributions of occupancy numbers under section Instructions c).

Social Decision Schemes applied:

Primary SDS: Median

Secondary SDS: Any of the two with equal probability, if there are two 'median' bets.

The group decision will fall on the median bet (or one of the median bets for 4-person groups).

Relevant Hypotheses:

The SDS Median, as described above, will predict group decisions correctly. No differences are expected between payoff conditions.

3.2 Results and discussion for Experiment II

In accord with the results of Experiment I we expected the Median decision rule to apply for bets with equal expected value (see hypotheses in box 3.1).

Table II/1: Frequency analysis of *Polarization* (from the Arithmetic Mean in direction of the Median) ^{+) ++)}

	Polarization		
	positive	o	negative
Arithmetic Mean	42	-	16
Median	4	38	16

Sign test for Arithm. Mean, 42 : 16 cases, $p < .01$

Sign test for Median, 4 : 16 cases, $p < .01$

^{+))} The analysis for this experiment is only conducted for frequencies. An analysis in terms of deviations does not provide more information since 56 out of 58 shifts are exactly one unit.

⁺⁺⁾ Since most shifts were risky, results for a risky shift analysis are almost identical with the polarization analysis (43 : 15 and 18 : 2 respectively).

Again there was a strong polarization (and shift also) from the arithmetic mean (see table II/1) but contrary to our foregoing results there was still polarization after the application of the Median decision rule. Although 47 of 58 group decisions fall exactly on (or one of) the median bet(s), polarization is significant; indicating that there is a systematic deviation from the median.

A closer look at the data reveals that the negative polarization from the median is primarily due to risk shift (18 : 2). The polarization is positive from the arithmetic mean (42 : 16) and negative from the median (4 : 16). This outcome is due to the fact that when a risk shift was expected from the arithmetic mean in direction of the median it occurred (group types II and III), whereas when a conservative shift was expected from the arithmetic mean in direction of the median it did not occur (group types I and IV). For the arithmetic mean taken as standard, this results in a positive polarization, for the median as standard it results in a negative polarization. In trying to explain these results, we suggest two alternative hypotheses:

1. *Money Scale Hypothesis*: When deciding on the bets, Ss are oriented on the money scale rather than the probability scale. In addition, the assumption underlying the median model of BLACK (1948) and COOMBS (1954) (that only the rankorder of alternatives is important for group decision, not the distances) is incorrect. In the case of Ss deciding primarily on the money scale with distances between alternatives as an important factor, it might be expected that - because of the rapid increase in money value - the arguments for bets with higher money values (and lower probabilities) would be more convincing in the group decision (see Box II, Instructions, Tasks, and Procedures, b)).
2. *Violation of Single Peakedness Hypothesis*: The basic social decision rule (majority voting over alternatives) is correct, but the condition of single peakedness of individual preference distributions is violated. Additionally, we have to assume that the violations do not wash out on the average but cause systematic deviations from the median decision rule.

Although we found in our studies that 86 % (263 of 305 cases) of the individual preference orders were single peaked, we have to expect theoretically that 48 % of the *groups* would have at least one individual violation of the condition of single peakedness (that would amount to 28 out of 58 cases). If the single peakedness had a systematic effect on the group decisions (e.g., towards risk), then 48 % of the groups were potentially affected.

Both alternative explanations shall be tested in the following experiment.

4. *Experiment III: Group Decision among bets of equal expected value, with two different types of bets (CROTT, SCHERMER, ZUBER, in prep.)*

4.1 *Brief description of the experimental conditions*

Subjects and conditions of participation:

Ss were 280 male students of law and economy. Each S got 5 DM for participation. Ss were informed that they could possibly additionally win a considerable amount of money. For each session, between 15 and 25 Ss were invited.

Instructions and Tasks:

- a) Explanation of the meaning of probabilities in terms of relative frequencies and the meaning of expected value as a long run expectation. Explanation of the meaning of individual preferences among bets with equal expected value (for the bets see table III/1). Ss were informed in advance that they have to discuss their preferences with others and have to come to a collective decision.
- b) Measurement of initial individual preferences by rankordering the six bets.

Procedure:

The Ss were assigned to 5-person groups with one of the following basic patterns of occupancy numbers:

- aa) 2 1 1 1 ba) 2 2 1 c) 2 1 2
 ab) 1 1 1 2 (mirror image) bb) 1 2 2 (mirror image)

Examples:

Bet.No.					
1	2	3	4	5	6
0	2	1	0	1	1
1	0	0	1	1	2

Bet.No.					
1	2	3	4	5	6
0	2	2	0	1	0
0	1	2	0	2	0

Bet.No.					
1	2	3	4	5	6
0	2	1	0	0	2
0	2	1	0	0	2

- c) Group discussion and decision (after 15 minutes).
- d) Second measurement of individual preferences: After the group decision,

individual preferences were measured a second time and the bet was played once.

- e) Repetitions (two times) of the procedure with set II and III (see table III/1).
- f) Playing out the bets: The chosen bet was played out once for the whole group and the amount of money of the bet or nothing was paid to the Ss depending on the results of the lottery.

Experimental Factors:

- a) The five types of distributions of occupancy numbers.
- b) Conditions P and DM for the structure of bets. Under condition P bets have equal probability (and consequently unequal money distances). Under condition DM bets have equal money distances (but unequal probability distances).

Social Decision Schemes applied:

For both conditions, P and DM, we try out the following SDS:

Primary SDS:

1. Arithmetic Mean
2. Median
3. Full Paired Comparison
4. Reduced Paired Comparison

Secondary SDS:

- Median bet in the range in case of intransitivities.
- Median bet in the range in case of intransitivities.

Example for SDS 3. and 4.:

Primary SDS: $\underline{1} > 2 < \underline{3} > 4 < \underline{5} > 6$

Secondary SDS: 3

Relevant Hypotheses:

There is no polarization in any of the conditions after application of the appropriate Social Decision Scheme.

4.2 Results and discussion for Experiment III

In order to make the investigation of the two hypotheses possible, the bets for Experiment III under condition P were constructed with equal probability distances (as before in Experiment II) and with equal money distances under condition DM (see table III/1). In addition, the complete rankorder of individual preferences over the alternatives was taken from each individual before and after the group decision.

The social decision scheme assumed to be correct according to hypothesis 2 for groups without an 'Absolute Majority' (see results and discussion for Experiment II) was the majority decision in paired comparison of alternatives. Besides the Arithmetic Mean and the Median, two variants of the Majority Decision by Paired Comparison of alternatives have been investigated. The

Table III/1: The *three sets of bets* used in the experiment under condition P and DM

	Set I:	Set II:	Set III:
P-con- dition	0.25/18.00	0.25/12.00	0.25/6.00
	0.35/12.80	0.35/ 8.60	0.35/4.30
	0.45/ 9.90	0.45/ 6.70	0.45/3.30
	0.55/ 8.20	0.55/ 5.50	0.55/2.70
	0.65/ 6.90	0.65/ 4.60	0.65/2.30
	0.75/ 6.00	0.75/ 4.00	0.75/2.00
DM-con- dition	0.24/18.50	0.26/11.50	0.33/4.50
	0.28/16.00	0.30/10.00	0.38/4.00
	0.33/13.50	0.35/ 8.50	0.43/3.50
	0.41/11.00	0.43/ 7.00	0.50/3.00
	0.53/ 8.50	0.55/ 5.50	0.60/2.50
	0.75/ 6.00	0.75/ 4.00	0.75/2.00

Table III/2: Decision Scheme 'Absolute Majority' for groups in Set II and III

Absolute Majority Prediction correct	Absolute Majority Prediction incorrect
40	12 ⁺⁾

⁺⁾ It doesn't make sense here to analyse the 12 cases according to risky shift and polarization. Only in two of the 12 cases risky shift *and* positive polarization are possible. The remaining 10 cases allow for cautious shift or negative polarization only, because the 'Absolute Majority' already predicts the extremest bet.

first one is called 'Full Paired Comparison/Median Bet' and the second one 'Reduced Paired Comparison/Median Bet' (see box). Full Paired Comparison means that all alternatives are compared in order to predict the group decision. For the Reduced Paired Comparison only the neighboured bets are compared successively, that is bet 1 vs. 2, bet 2 vs. 3 and so on. In both bases (but especially with the Reduced Paired Comparison) the resulting distribution of the group preference order might not be transitive. As a secondary decision rule we take the median bet of an intransitive sequence as the best estimation (see example in box). If hypothesis 1 was true, we

should get differences in the P and DM condition even after applying the decision rule considered appropriate; according to hypothesis 2 (i.e., under the P condition) there should remain polarization, but not under the DM condition.

Table III/3: Frequencies of *groups* without violations of single peakedness (or irrelevant violations) and *groups* containing at least one member with violations of single peakedness for condition P and DM separately.

	Groups without violations of single peakedness (or irrelevant violations)		Groups with relevant violations of single peakedness		
P	22	(26.5)	20	(15.5)	42
DM	31	(26.5)	11	(15.5)	42
	53		31		

$$\chi^2 = 4.14, df = 1, p < .05, \text{descriptive } ^+)$$

^{+) 29 groups of set II and III which consist of same Ss as groups from set I are included in the table, so frequencies are partly dependent.}

In table III/3 we can compute that the percentage of groups with at least one member with at least one relevant violation of the single peakedness is 37 %. A violation is considered relevant when the Full Paired Comparison of all alternatives by majority voting leads to predicted group decision different from the median. We can also see from table III/3 that there are more groups with relevant violations under the p than under the DM condition.

Polarization and shift have first been analysed in terms of frequencies (comparison of number of cases with positive vs. number of cases with negative polarization and number of cases with risky or with conservative shift). Results of that analysis are given in tables III/3 and III/4. Table III/5 shows the polarization results of the analysis of deviations.

Table III/4: *Frequencies of Polarization (to the Median) under condition P*

	positive	no	negative	χ^2 -test ($p \leq$)
Arithmetic Mean	20	0	9	.10
Median	6	14	9	.70
Full Paired Comp.	8	15	6	.80
Reduced Paired Comp.	11	14	4	.20

Frequencies of Polarization (to the Median) under condition DM ⁺)

	positive	no	negative	χ^2 -test ($p \leq$)
Arithmetic Mean	34	0	8	.001
Median	6	28	8	.80
Full Paired Comp.	11	21	10	.90
Reduced Paired Comp.	10	22	10	.99

⁺) Regarding shift there are significant results for all four models (.0001, .01, .05, .20) under the P condition and for the 'Arithmetic Mean' under the DM condition (.05).

To provide an overview in table III/6 the remaining probabilities for all statistical tests are collected. The values above the diagonal refer to the frequency analyses the values below to the deviation analyses. - It can be seen from column 1 that the group decision is significantly different from the Arithmetic Mean, this is true for frequency- and for deviation analysis, for risky shift and polarization, for condition P and DM and for both conditions together. Among the 12 values there are only two exceptions for polarization deviations under condition P (with $p < .10$ and $p < .30$). ¹⁾ The

¹⁾ Regard that the Arithmetic Mean and the Median are not the hypothesized Decision Schemes, so we take the 5 % level of significance in these cases, and the 20 % level for the following tests of the Full Paired Comparison and the Reduced Paired Comparison.

Table III/5: *Deviation Analysis of Polarization (to the Median): Deviation from the Social Decision Scheme prediction (group decision minus SDS prediction) Set I, II, and III 'Average' +) ++)*

1 st SDS	Arithmetic Mean	Median	Full Paired Comp. +++)	Reduced Paired Comp. +++)
2 nd SDS			Median Bet	Median Bet
P	n/df	25/24	25/24	25/24
	$\bar{x}$	0.027	-0.016	0.020
	s	0.112	0.121	0.118
	t	1.21 ^{.30}	-0.66 ^{.60}	0.85 ^{.50}
DM	n/df	27/26	27/26	27/26
	$\bar{x}$	0.032	-0.002	-0.001
	s	0.045	0.036	0.054
	t	3.68 ^{.01}	-0.24 ^{.90}	-0.08 ^{.90}
P+ DM	n/df	52/50	52/50	52/50
	$\bar{x}$	0.030	-0.009	0.009
	s	0.084	0.088	0.091
	t	2.58 ^{.02}	-0.74 ^{.50}	0.71 ^{.50}

+) Since Ss of the groups in Set II and III which correspond to the occupancy numbers as described in Box 4.1. are identical with the Ss in Set I, we averaged the deviations from the predicted values for Set I, II, and III in cases in which there was more than one set with critical occupation numbers.

++) Considering only risky shift 'Arithmetic Mean' shows significant deviations for P, DM and P + DM (with $p < .0001$, $< .01$, $< .001$, respectively). The median model shows significant deviations for P and P + DM ($< .01$, $< .01$) whereas for DM there is no risky shift. For the model 'Full Paired Comparison' the probability is below .20 only for P + DM together. For the 'Reduced Paired Comparison' model the probabilities do not go below .20.

+++ As 'Full Paired Comparison' and 'Reduced Paired Comparison' sometimes predict the most extreme bet, risky shift and positive polarization were not always possible under those models. Therefore a conservative correction for floor/ceiling effects was applied by enlarging the risky shift and positive polarization one standard deviation. However, this procedure had very little effects on the $\bar{x}$ and t-test values, so that floor/ceiling effects apparently can be neglected in the given experimental material.

Table III/6: Summary table of significances of frequency- and deviation analysis for risky shift and polarization (see also tables III/4 and III/5). Frequency significances above, deviation significances below diagonal

		Arithmetic Mean	Median	Full P. C. median bet	Reduced P. C. median bet
P	risky shift	.001	.01	.05	.20
	polarization	.001	.01	.30	.30
DM	risky shift	.10	.70	.80	.20
	polarization	.30	.60	.80	.50
P + DM	risky shift	.05	.50	.70	.60
	polarization	.01	.70	.70	.50
P	risky shift	.001	.01	.10	.70
	polarization	.001	.01	.20	.50
DM	risky shift	.001	.01	.80	.40
	polarization	.01	.90	.80	.50
P + DM	risky shift	.001	.01	.10	.70
	polarization	.001	.01	.20	.50

Median (column 2) shows some significant deviations regarding risky shift especially under the P condition. Polarization from the Median is not significant in any case. Also for the Full Paired Comparison/Median Bet there are still some significant results for risky shift under the P condition and both conditions taken together ($p < .05$ and $< .10$). For the Reduced Paired Comparison/Median Bet there are only marginal significances left for risky shift and polarization in frequency analysis under P, that vanish when P and DM are taken together.

These results demonstrate that the violation of single peakedness had a systematic effect on the group decisions. When taking individual character-

istics into account by using the Paired Comparison Model, however, the risky shift and polarization reduce considerably under any condition, P or DM. Results are also summarized in Figure III/1 which gives a graphical representation of the reductions in risky shift and polarization by application of appropriate decision rules. By the Reduced Paired Comparison/Median Bet risky shift is reduced to about 17 % of the original amount, polarization to approximately 30 %.

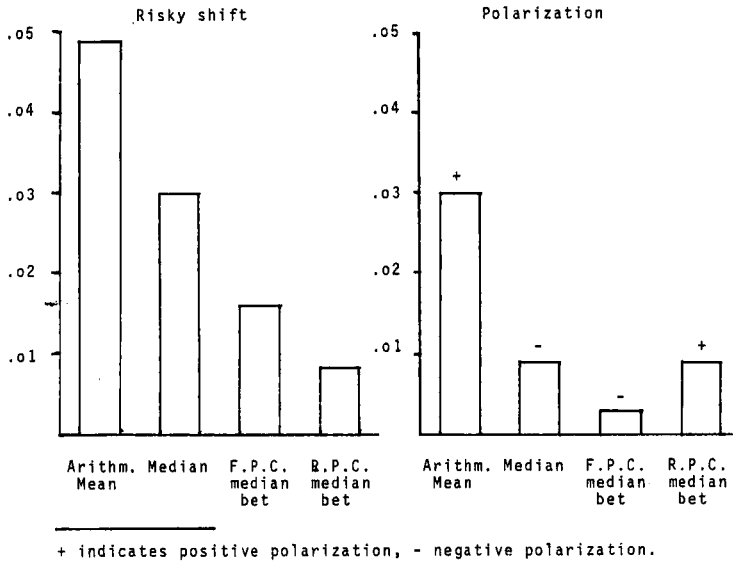


Figure III/1: Graphical representation of risky shift and polarization effects for P and DM

We conclude risky shift and polarization can be explained by the effect of the decision rule. Admitting that there is still some risky shift and polarization left (sometime positive, sometime negative) nevertheless risky shift and polarization are reduced considerably by applying the appropriate decision rule.

Hypothesis 2 is supported rather than hypothesis 1. This corresponds well with Experiment I, where there was no systematic deviation from appropriate decision rule although the distribution of estimations within groups tended to be fairly skewed.

5.1 Summary and conclusions for Experiments I to III

In all three experiments (in total 178 systematically varied groups) we found significant polarization from the arithmetic mean in direction of the median. In experiment II and III risky shift from the arithmetic mean was significant as well. In experiment I Ss had to estimate a number of points on a slide briefly exposed to them. Then they had to engage in a collective estimation. Since single peakedness of individual preference functions appears to be a natural property of the task (in the sense of BLACK), the median is the appropriate *primary* social decision rule. Results show that when depending on the motivational conditions, the two variants of the median model are applied to the data, polarization vanishes almost completely.

In experiment II, in which Ss had to decide collectively upon bets of equal expected value, the median did not reduce risky shift and polarization to an insignificant level. Even though 47 of 58 decisions were on the (or *one* of the) median values, the testing procedure revealed significant (negative) polarization and risky shift. Regarding to forgoing results, experiment III was designed to test with a similar task the following alternative hypotheses:

1. The assumption that only rankorders are important for group decisions (not distances) is wrong. W.o.w. the assumption underlying the BLACK (median) model is wrong, with regard to the actual behavior of individuals in a group decision process.
2. The basic assumption of the Black model is correct, but the formulation of the decision has to take into account violations of the single peakedness condition. The two decision rules defined in addition to the median were Full Paired Comparison and Reduced Paired Comparison.

Other decision rules like Proportionality, Equiprobability, and Plurality were also tested. The results are not reported here, however, since their predictions are less reliable than any of the other three decision rules (Median, Full Paired Comparison, Reduced Paired Comparison). Moreover, they are sometimes even less precise in prediction than the Arithmetic Mean.

Application of the decision rules Full and Reduced Paired Comparison often leads to more risky predictions for the group decision, indicating that the violations of the single peakedness condition do not wash out in the average regarding risky and conservative shifts. This supports the hypothesis that the observed risky shift that goes beyond the median model is due to violations of the single peakedness condition.

The results indicate that the basic ideas of majority voting over alternatives accounts for most of the biases called *risky shift* or *polarization*.

We do not claim that genuine risky/cautious shift and polarization are non-existent. Nevertheless we indicate that a more careful analysis of the collective decision processes should precede a shift or polarization analysis in order to be able to test the existence of true shift/polarization effects.

References

- Arrow, K.J.: 1951. Social choice and individual values. New York: Wiley
- Black, D.: 1948. On the rationale of group decision making. In: Journal of Political Economy 56, 23-34
- Coombs, C.H.: 1954. Social choice and strength of preference. In: R.N. Thrall/ C.H. Coombs & R.L. Davis (eds.), Decision processes. New York, 69-86
- Crott, H.W.: 1964. Experimentelle Untersuchung eines Modells für das Gruppenentscheidungsverhalten in Verhandlungssituationen. Zulassungsarbeit am Psychologischen Institut der Universität Hamburg
- Crott, H.W.: 1965. Gruppenentscheidungen unter verschiedenen Motivationsbedingungen. In: Archiv für die gesamte Psychologie 117, 113-130
- Crott, H.W./Schermer, Th./Knoll, C. & Thoene, B.: Experimental analysis of social decision schemes and risky shift in group decisions among bets. In prep.
- Crott, H.W./Schermer, Th. & Zuber, J.A.: Experimental analysis of social decision schemes in various group decision tasks. In prep.
- Davis, J.H.: 1973. Group decision and social interaction. A theory of social decision schemes. In: Psychological Review 80, 97-125
- Hoffman, P.J./Festinger, L. & Lawrence, D.R.: 1954. Tendencies toward group comparability in competitive game behavior. In: Human Relations 7,

141-159

- Lamm, H. & Myers, D.G.: 1978. Group-induced polarization of attitudes and behavior. In: L. Berkowitz (ed.), *Advances in experimental social psychology*, Vol. XI. New York: Academic Press
- Moscovici, S. & Zavalloni, M.: 1969. The group as a polarizer of attitudes. In: *Journal of Personality and Social Psychology* 12, 125-135
- Myers, D.G.: 1982. Polarizing effects of social interaction. In: H. Brandstätter/J. Davis & G. Stocker-Kreichgauer, *Group decision making*. London: Academic Press
- Stoner, J.A.F.: 1961. A comparison of individual and group decision including risk. Unpublished diploma paper, School of Industrial Management, M.I.T.

**THE DEPENDENCE OF COGNITION AND BEHAVIOR
IN 2x2 MATRICE GAMES**

ULRICH SCHULZ

THEO MAY

*Universität Bielefeld
Abteilung für Angewandte und Experimentelle Psychologie
F.R. Germany*

In small group research experimental games are studied which have the following simple characteristics:

The game is played by two subjects. Each subject has two alternative choices (strategies). The game is played without bargaining or communication. Therefore each player's decision is made independently from the other subject. The outcome of the game consists in a pair of choices (strategies). Each outcome has a utility for each subject, usually measured in money (payoff). In most experiments the payoff is displayed in matrix form:

		subject 2	
		strategy I	strategy II
subject 1	strategy I	payoff of S1 for outcome (I,I)	payoff of S1 for outcome (I,II)
	strategy II	payoff of S1 for outcome (II,I)	payoff of S1 for outcome (II,II)

Both subjects know their own and their partner's payoff.

In conflict games there exists no strict competition between the subjects as the profit to one subject causes no loss for the other. The payoff structure is designed in such a way that choices, based only on the own utility, result with certainty in low payoff decisions. If the other's payoff is also taken into consideration, this may result in uncertain outcomes: It is possible that reasonable payoffs for both subjects can be reached but a low payoff is also possible. In such conflicts the principle of individual rationality no longer provides reasonable solutions. An. RAPOPORT argued that the study of gambling behavior might provide new indications leading to a solution concept for these games.

1. Prisoner's Dilemma Game and Chicken Game

Two-person games are considered in which each person i ($i = 1, 2$) has two pure strategies called C and D. The game is played without communication. Player i 's payoffs are denoted by $a_i(X, Y)$, where X and Y symbolize the respective choices of the player and of his opponent. Prisoner's Dilemma Games (PDG) are defined, if we only consider restricted and symmetric games by the inequalities:

$$\begin{aligned}
 & a_1(D, C) > a_1(C, C) > a_1(D, D) > a_1(C, D) \\
 (1) \quad & 2a_1(C, C) > a_1(D, C) + a_1(C, D) \\
 & a_2(X, X) = a_1(X, X) \text{ for } X = C, D \\
 & a_2(X, Y) = a_1(Y, X) \text{ for } X \neq Y \quad X, Y = C, D
 \end{aligned}$$

The symmetric and restricted Chicken Game (CG) is defined by the inequality

$$a_1(D, C) > a_1(C, C) > a_1(C, D) > a_1(D, D)$$

In the CG the order of the outcomes D,D and C,D are simply reserved. All other definitions remain the same.

Both games belong to the class of conflict games. Conflict arises in these games from the fact that at least the payoff of one player for equilibrium points is lower than the payoff for the strategy combination C,C.

2. *Psychological Research in Experimental Matrice Games*

Psychological researches have mainly focused on the topics of conflict resolution, social motivation, and information processing in iterated games (e.g. KRIVOHNAVY, 1974). Beside studying the player's choice behavior, psychologists soon became very interested in the subject's prediction of the other player's behavior and the accuracy of this prediction. The main results of this research can be summarized as follows:

Additional registration of prediction does not influence decision and gaming behavior (TEDESCHI et al., 1968; HALPIN & PILISUK, 1970; SCHULZ & MAY, 1982).

Changes in predictions in the course of the gaming sequence show that players process information about the other player's gaming behavior. (SERMAT, 1964; BIXENSTINE & BLUNDELL, 1966; PYLYSHYN et al., 1966; TEDESCHI et al., 1968; HALPIN & PILISUK, 1970; REVENSTORFF et al., 1977; SCHULZ & HESSE, 1978; BLUDSZUWEIT & KAUK, 1976; SCHULZ & JONAS, 1982).

There is a minor group of players whose predictions become increasingly more accurate in the course of gaming. These subjects are called realists. The majority of subjects is not realistic. (PYLYSHYN et al., 1966; TEDESCHI et al., 1968; BLUDSZUWEIT & KAUK, 1976; BARTH, 1979; FRINDTE, 1979).

The combination of prediction and strategy choice in each game of the sequence may be interpreted as a motivational state (e.g. trust, risk etc.). (LOOMIS, 1959; DEUTSCH, 1960; SOLOMON, 1960; TERHUNE, 1968; REVENSTORFF et al., 1977; SCHULZ & HESSE, 1978).

The subject's information processing causes changes in the frequencies of the motivational states in the gaming sequence. (HALPIN & PILISUK, 1970; SCHULZ & HESSE, 1978).

The majority of realists have a noncooperative motivation. They take their predictions into consideration and try to maximize their utility. (BLUDSZUWEIT & KAUK, 1976; FRINDTE, 1979; BARTH, 1979).

Because of these results, and as an attempt to explain the phenomenon of realism, SCHULZ & MAY (1982) performed PDG-experiments where subjects had to indicate their prediction, their subjective certainty of the prediction, and their choice. The authors argued that for realists, subjective certainty of prediction should highly correlate with correct predictions. In the experiments the subject's subjective certainty was indicated by binary forced choice

between the two states "certain" and "uncertain".

3. *A Model for Choice Behavior, Prediction and Subjective Certainty*

SCHULZ & MAY (1982) proposed a model for the choice, the prediction and it's certainty, in sequences of PDG which is presented here along with a more specified psychological interpretation of some processes.

The subject's process with regard to the other player's choice is the estimation process for the unknown probability π that the other will cooperate. General experience and information gained during the preceding games specifies some probability distribution of the unknown parameter π as indicated in figure 1. (Similar models can be found as prior distribution in Bayesian statistics (e.g. SCHAEFER, 1976; WINKLER, 1967).)

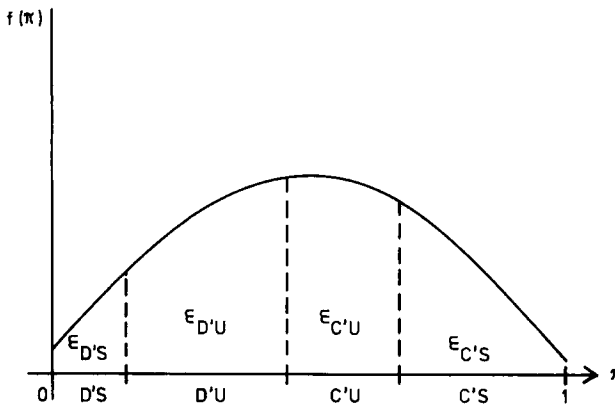


Fig. 1

If we denote the subject's predictions by C' and D' and the states "certain" by S and "uncertain" by U the pairings of prediction and certainty can be interpreted as labels for ordered intervals on the axis of the parameter π as indicated in figure 1.

If the subject has to predict the others choice, and it's certainty, he/she

is sampling according to the probability distribution $f(\pi)$ a value of π^* . The subject looks for the interval which contains π^* and reports the label of this interval. Interval boundaries are considered to be fixed. Observable probabilities of the intervals can be calculated by summation or integration. The probability distribution is considered to vary during the sequence of the games. The probabilities of the respective prediction-certainty stated conditional on the gambling history of the process are denoted by $\epsilon_{D'S,n}, \epsilon_{D'U,n}, \epsilon_{C'U,n}, \epsilon_{C'S,n}$ (n = number of trials). Theoretically these probabilities can be calculated by integration of the actual distribution $f(\pi)$. The model assumes that f changes in such a way that observable probabilities ϵ are changed by learning operators:

denote by

$$(3a) \quad \epsilon_{X,n+1} = g_{X,X}(\epsilon_{X,n}) = \frac{\alpha_X \epsilon_{X,n}^{1/\alpha_X}}{\alpha_X \epsilon_{X,n}^{1/\alpha_X} + 1 - \epsilon_{X,n}}$$

$X = C'S, C'U, D'U, D'S;$
 $\alpha_X > 0.$

$$(3b) \quad \epsilon_{Y,n+1} = g_{X,Y}(\epsilon_{Y,n}) = \frac{\epsilon_{Y,n}}{\alpha_X \epsilon_{X,n}^{1/\alpha_X} + 1 - \epsilon_{X,n}}$$

$X \neq Y \text{ and } X,Y = C'S, C'U, D'U, D'S.$

These two equations specify the learning operators applied to the probabilities ϵ . Figure 2 shows the graph of the function $g_{X,X}$. These operators are simple one parametric noncommutative operators.

The player's prediction, and his partner's choice in trial n of the game sequence are indicated in the first and second column of table 1. The third column indicates which probability will be transformed after observation of the specified prediction-choice combination of column 1 and 2 by the operator $g_{X,X}$. The last column specifies the parameter used in $g_{X,X}$. All other probabilities $\epsilon_{Y,n}$ are transformed by $g_{X,Y}$ using the appropriate

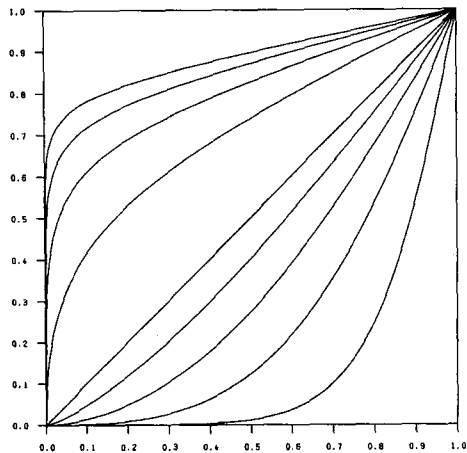


Fig. 2

Table 1 contains the reinforcement schedule of the model:

Table 1 (Reinforcement schedule)

prediction in trial n	choice by alter in trial n	X for which $\epsilon_{X,n+1} = g_{X,X}(\epsilon_n)$	learning parameter
C'	C	C'S	$\alpha_{C'S}$
C'	D	C'U	$\alpha_{C'U}$
D'	C	D'U	$\alpha_{D'U}$
D'	D	D'S	$\alpha_{D'S}$

parameter α .

All the information processing assumptions of the model are now specified by equations. For each player we introduced four starting probabilities $\epsilon_{X,1}$, $X = C'S, C'U, D'U, D'S$, three of which have to be estimated. Further parameters are the four learning parameters of table 1, which also have to be estimated.

The model assumes that in each game the first stage in each player's decision and choice process is described sampling of a prediction-certainty state.

The model further assumes that given the prediction-certainty state, players have fixed propensities to choose between C and D. These propensities are independent of the history of the gambling process and of the number of trials. They are measured by four conditional probabilities:

$$(4) \quad \begin{aligned} \xi_{C'S} &= P(C | C'S) & \xi_{D'U} &= P(C | D'U) \\ \xi_{C'U} &= P(C | C'U) & \xi_{D'S} &= P(C | D'S) \end{aligned}$$

The simple notation is justified by the above mentioned independence assumptions. The propensities ξ are four additional parameters introduced for each player, so that each player is characterized by 11 unknown parameters.

4. *Validation of the Model for PDG and CG*

a) Experiments

Experiments were performed with 3 PDG matrices and 3 CG matrices. SCHULZ & MAY (1982) describes the experimental procedure in detail. Subjects were male students, but not psychology students. Subjects were payed for their participation. Their remuneration depended on the results of the experimental games.

Table 2 indicates the matrices used, the number of trials in the sequence and the number of players.

Table 2

Code of game	matrix		number of trials	number of subjects
	C	D		
CG0	C	12 -2	45	30
	D	22 -6		
CG1	C	4 -2	50	20
	D	9 -8		
CG2	C	4 -2	50	18
	D	9 -6		
PDG1	C	4 -3	50	20
	D	9 -2		
PDG2	C	4 -3	50	22
	D	9 +1		
PDG0	C	8 -6	45	30
	D	18 4		

b) Parameter Estimation and Model Control

In experimental games, research into detailed processes only seems to be successful if interindividual differences (due to different motives, different experiences and different information processing) are taken into consideration. In the absence of better methods, the parameters of individual subjects are estimated by Maximum-Likelihood-Methods using data provided by individual pairs of players. It should be mentioned that the parameters ξ are estimated by conditional relative frequencies (for more detailed information see SCHULZ & MAY, 1982).

Model tests are also performed for individual pairs of players. Indices, such as frequency of bilateral cooperation in the first or second half of the

sequence of games, can be computed from empirical data. By simulation of the model using the estimated parameters it is possible to decide whether the observed index values are extreme and unlikely if the model is valid ($\alpha = 0.05$). If the indices are extreme and unlikely the model is not able to describe these aspects of empirical data. (For more details of the model tests see SCHULZ & MAY, 1982). As in SCHULZ & MAY (1982) 48 indices per pair of players were tested using $\alpha = 0.05$ per test. If the model is valid, two or three violations can be expected. Therefore we decided to consider that more than three model violations produce serious discrepancy between data and model prediction.

Table 3 shows the number of serious model violations for the six games.

Table 3 Model rejections

Game	CG0	CG1	CG2	PDG1	PDG2	PDG0
number of pairs with model reject.	0	1	0	3	1	0

For PDG, the model was rejected in 11 % of all cases. For CG, the model was rejected in 3 % of the cases.

We conclude that the model can be used for the description of decision and choice behavior in sequential PDG and CG. (Detailed results of the model tests for PDG are reported in SCHULZ & MAY, 1982, for CG in SCHULZ & MAY, 1983.) The pairs of players for which the model was rejected were excluded from further analysis.

5. *The Influence of Certainty on Choices*

The model stated an influence of subjective certainty on choice behavior. Earlier investigations of choice behavior in experimental games have shown that simple models involving only choice and prediction can be used to describe the main features of decision and choice processes in PDG (SCHULZ & JONAS, 1982; MAY, 1982). Therefore the question arises as to whether certainty of prediction is necessary for the description of decision and choice be-

havior. This question can be answered by examining the submodel which is characterized by

$$\xi_{C'S} = \xi_{C'U}$$

$$\xi_{D'S} = \xi_{D'U}$$

As described in SCHULZ & MAY (1982), the support of the submodel by data relative to the support of the general model can be demonstrated by considering the differences of the doubled log-likelihood-values of the submodel and the general model. Because we know nothing about the distribution of the likelihood ratios we cannot perform the usual chi-square test. Using a critical difference of 9.5 for the double of log-likelihood differences (corresponding to $\alpha = 0.05$ in a chi-squared distribution with 4 degrees of freedom), the general model was distinctly better supported by experimental data than the submodel (in 24 % of the CG-cases and in 13 % of PDG-cases).

More detailed inspection shows that the influence of certainty may depend on the payoff matrix. In 40 % of the CGO cases the general model was distinctly better supported. Our conclusion is that the submodel should not be used for CG.

6. *Influence of Payoff Matrix on Model Parameters*

Descriptive comparison of behavior in PDG and CG has been reported by several authors (RAPOPORT & CHAMMAH, 1969; SERMAT, 1967; SCHLENKER & GOLDMAN, 1978). The general result is that the different character of the conflicts in CG and PDG leads to higher cooperation in CG.

The next step, in which behavior and experimental games is analysed by mathematical models, consists of an attempt to investigate the relations between model parameters and situational parameters, particularly the payoff matrix of the game. Our attempt concentrates on the parameters ξ .

Theoretical and experimental work in experimental games (e.g. DEUTSCH, 1960; DEUTSCH, 1973; McCLINTOCK, 1972; McCRIMMON & MESSICK, 1976; KELLEY & THIBAUT, 1978) has shown that social motives heavily influence the subject's evaluation

of the payoff matrix. The majority of these authors postulate that the subjective utility of the game is some simple algebraic function of the payoff matrix, such as the sum or the difference of the payoffs. This function depends on the subject's social motivation. This subjective transformation may lead to a conflict structure of the subjective utilities that differs from the previous payoff matrix. Therefore (when considering different motivations) it may be difficult to derive general results for different games without knowledge of the subject's motivational tendencies.

Consequently experiments with special payoff matrices were designed in order to demonstrate some effects of the payoff matrix on the propensities of subjects which would be valid under fairly weak assumptions about the subjective utility structure of the payoff matrices.

For some strategy combination of both players Z ($Z = CC, CD, DC, DD$), let x_Z denote the player's own payoff and y_Z the payoff of his partner. Then (according to LUCE & SUPPES, 1965) we may assume the existence of a subjective strength for each payoff combination (x, y) denoted by $v(x, y)$. It is assumed that strength is some strict monotonically increasing transformation of utility. The main property of strength is its positivity $v(x, y) > 0$. We assume three further properties of strength:

Translation invariance of ratios:

$$(5) \quad \frac{v(x, y)}{v(x^*, y^*)} = \frac{v(x + d, y + d)}{v(x^* + d, y^* + d)}$$

Monotonicity for equal payoffs:

$$(6) \quad v(x, x) > v(y, y) \text{ for } x > y$$

Monotonicity on constant payoff for the other:

$$(7) \quad v(x, z) > v(y, z) \text{ for } z > x > y$$

The next step is to find a relation between the probabilities ξ and the corresponding strengths.

Going back to figure 1 we assume that each interval has a characteristic value of subjective probability π_Z , $Z = C'S, C'U, D'U, D'S$ (e.g. the midpoints for the interval).

$$\pi_{C'S} \geq \pi_{C'U} \geq \pi_{D'U} \geq \pi_{D'S}$$

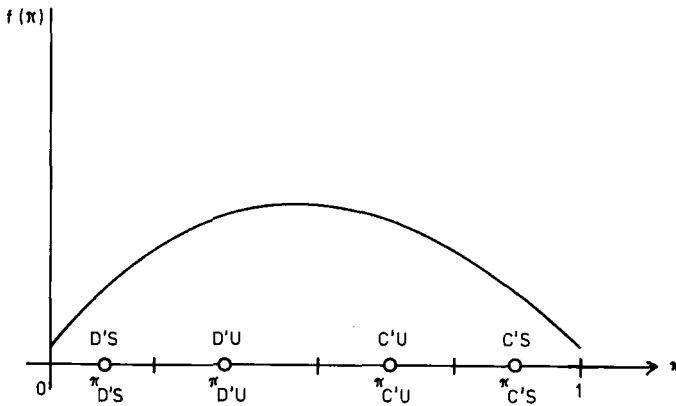


Fig. 3

We further assume that the probabilities of choices, conditional on the prediction-uncertainty states, are influenced by the characteristic values π_Z according to the principle of strict expected utility introduced by LUCE & SUPPES (1965).

$$\xi_Z = \frac{\pi_Z v(x_{CC}, y_{CC}) + (1-\pi_Z) v(x_{CD}, y_{CD})}{\pi_Z (v(x_{CC}, y_{CC}) + v(x_{DC}, y_{DC})) + (1-\pi_Z) (v(x_{CD}, y_{CD}) + v(x_{DD}, y_{DD}))} \quad (8)$$

for $Z = C'S, C'U, D'U, D'S$

Through the assumptions (5)-(7) we are now able to derive some special predictions about the parameters ξ_Z for the matrices displayed in table 4.

Table 4

Experiment		$\xi_{C'S}$	$\xi_{C'U}$	$\xi_{D'U}$	$\xi_{D'S}$
CGO	30	.86 (1.0)	.53 (1.0)	.75 (1.0)	.60 (1.0)
CG1	18	1.0 (.97)	.71 (.75)	1.0 (.80)	.75 (1.0)
CG2	18	1.0 (.89)	.93 (1.0)	.56 (1.0)	.33 (1.0)
PGG1	14	.64 (1.00)	.67 (1.00)	.20 (1.00)	.17 (.40)
PGG2	20	1.0 (1.0)	.76 (1.0)	0.0 (.64)	.0 (.60)
PDGD	30	.89 (1.0)	.69 (1.0)	.15 (.83)	.16 (.89)

6.1 Comparison of CGO and PDGO

With assumption (8) parameters ξ_Z are invariant under translation of payoffs.

Therefore CGO and the CG have the same propensities with the following matrix:

		C	D	2
ξ_Z :	C	8/8	-6/18	
	D	18/-6	-10/-10	

For the outcomes, C,C; C,D; D,C this matrix has the same payoffs as PDGO. Applying assumptions (5) to the strength of outcome DD of this matrix we conclude:

$$\xi_{Z,CGO} \geq \xi_{Z,PDGO} \text{ for } Z = C'S, C'U, D'U, D'S$$

Since $\pi_{C'S} \approx 1$ and $\pi_{C'U} \gg 0$ is a plausible additional assumption, we may differentiate further: the difference between $\xi_{CS, CGO}$ and $\xi_{CS, PDGO}$

$$\xi_{CS,CGO} \approx \xi_{CS,PDGO}$$

The difference between $\xi_{CU, CGO}$ and $\xi_{CU, PDGO}$ should be moderate but the total influence of the outcome DD will be observed for the other two ξ_Z , because π_Z is then small, almost 0. Therefore we expect

$$\xi_{DU, CGO} > \xi_{DU,PDGO}$$

$$\xi_{DS, CGO} > \xi_{DS, PDGO}.$$

If we take independent samples of players from the same population for CGO we predict significant differences in the central tendencies of ξ_{DU} and ξ_{DS} for both games, and at most, slight differences for ξ_{CS} between these two games. Mann-Withney-tests for the two variables show significant differences only for ξ_{DU} and ξ_{DS} , and no significant difference for ξ_{CS} (the tests were performed on the 5%-level).

6.2 Comparison of CG1, CG2, PDG1, PDG2

CG1 and CG2 differ only in the payoffs of the DD outcome. According to (5) and variation π_Z we conclude

$$\xi_{Z,CG1} \geq \xi_{Z,CG2} \quad Z = C'S, \dots, D'S$$

$$\xi_{C'S, CG1} \approx \xi_{C'S, CG2}, \quad \xi_{D'U, CG1} > \xi_{D'U, CG2}$$

$$\xi_{D'S, CG1} > \xi_{D'S, CG2}$$

Using the same arguments for PDG1 and PDG2 produces

$$\xi_{Z, PDG1} \geq \xi_{Z, PDG2} \quad Z = C'S, \dots, D'S$$

$$\xi_{C'S, PDG1} \approx \xi_{C'S, PDG2} \quad \text{and} \quad \xi_{D'U, PDG1} > \xi_{D'U, PDG2}$$

$$\xi_{D'S, PDG1} > \xi_{D'S, PDG2}$$

Relations between CG2 and PDG1 can only be established for small values of π_Z , especially for $\pi_{D'U}$ and $\pi_{D'S}$ near 0. In this case $\xi_{D'U}$ and $\xi_{D'S}$ assumptions (7) and (6) it follows:

$$v(-2,9) > v(-3,9)$$

$$v(-2,-2) > v(-3,-3)$$

We therefore conclude

$$\xi_{D'U, CG2} > \xi_{D'U, PDG1}$$

$$\xi_{D'S, CG2} > \xi_{D'S, PDG1}$$

If we combine all separate hypotheses we expect no significant differences between CG1 and CG2, or between PDG1 and PDG2, in the variable $\xi_{C'S}$. Summarizing the inequalities for the parameters $\xi_{D'U}$ we expect decreasing central tendencies going from CG1, over CG2 and PDG1, to PDG2. The same hypotheses can be stated for the variable $\xi_{D'S}$. Median values of the probabilities are depicted in figure 5.

The last two hypotheses were tested by the Kruskal-Wallis-test. There were significant differences in the variables $\xi_{D'S}$ and $\xi_{D'U}$ ($\alpha = 0.05$). Figure 4 also demonstrates the monotonic decreasing tendency.

There was no significant difference between CG1 and CG2 in the variable $\xi_{C'S}$, but there was a significant difference in this variable between PDG1 and PDG2 ($\alpha = 0.05$).

7. Discussion

Prediction of the other's choice and its certainty can be integrated in a model for decision and choice behavior in PDG and CG. The model has two somewhat independent parts: information processing that only influences predictions and certainty, and the constant strategic propensities. The model can be used for PDG and CG because only a few model rejections can be stated. The analysis of the submodel shows that subjective uncertainty has to be considered at least in CG. To resolve this question within the system of classical statistics more data should be collected.

The interpretation of the four prediction-certainty-states as being ordered intervals with characteristic subjective probabilities can be used for the derivation of predictions for the propensities ξ , if weak conditions are fulfilled by the subjective strength of game matrices, and if the model of strict expected utility is valid. Though this theoretical derivation may not be the only explanation of the relations between the parameters, the proposed interpretation of the decision and choice process in terms of traditional theory is able to show that using fairly weak assumptions, interesting aspects of the parameters can be predicted. Perhaps it would be possible to formalize these assumptions and demonstrate their validity in more detailed experiments.

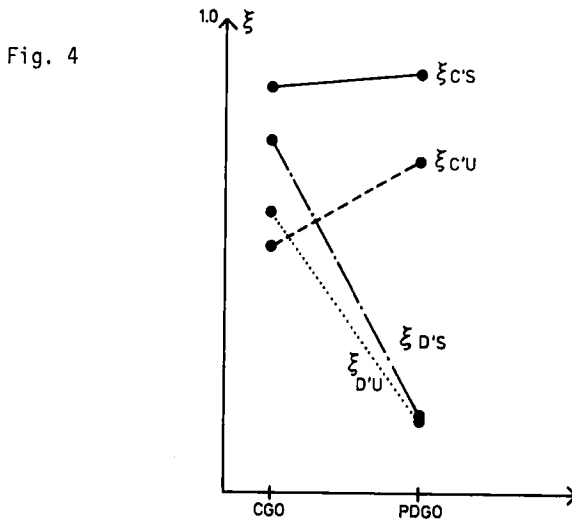
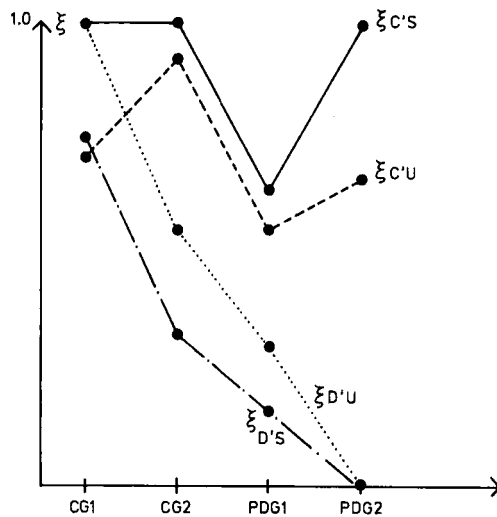


Fig. 5



References:

- Barth, W.: 1979. Faktorenstruktur verschiedener Nichtnullsummenspiele unter besonderer Berücksichtigung des PDG. In: *Wiss. Ztschr. Friedrich-Schiller-Univ. Jena, Ges.- und Sprachwiss.*, 28, 817-823.
- Bixenstine, V.E. & Blundell, H.: 1966. Control of choice exerted by structural factors in two-person, non-zero-sum games. In: *Journal of Conflict Resolution*, 4, 478-487.
- Bludszuweit, H. & Kauke, M.: 1976. Kooperatives Verhalten und realistische Partnererwartungen im Spielexperiment. In: *Zeitschrift für Psychologie*, 184, 157-170.
- Deutsch, M.: 1960. The effect of motivational orientation upon trust and suspicion. In: *Human Relations*, 13, 123-140.
- Deutsch, M.: 1973. *The resolution of conflict*. Yale University Press, New Haven.
- Frindte, F.: 1979. Erfahrungen mit einem neuen experimentellen Spiel. In: *Wiss. Ztschr. Friedrich-Schiller-Univ. Jena, Ges. u. Sprachwiss.*, 28, 6.
- Halpin, S.M. & Pilisuk, M.: 1970. Prediction and choice in prisoner's dilemma. In: *Behavioral Science*, 15, 141-153.
- Kelley, H.H. & Thibaut, J.W.: 1978. *Interpersonal relations: A theory of interdependence*. Wiley, New York.

- Krivohlavy, J.: 1974. Zwischenmenschliche Konflikte und experimentelle Spiele. Huber, Bern.
- Loomis, J.L.: 1959. Communication, the development of trust, and cooperative behavior. In: *Human Relations*, 12, 305-315.
- Luce, R.D. & Suppes, P.: 1965. Preference, utility and subjective probability. In: Luce, R.D., Bush, R.R. & Galanter, E.(Eds.): *Handbook of mathematical psychology*, Vo. III, Wiley, New York.
- May, T.: 1982. Einfluß der Situation auf das Verhalten in experimentellen Konfliktspielen. Paper presented on the 24. Tagung experimentell arbeitender Psychologen, Trier.
- McClintock, C.G.: 1972. Social Motivation - a set of propositions. In: *Behavioral Science*, 17, 438-454.
- McCrimmon, K.R. & Messick, D.M.: 1976. A framework for social motives. In: *Behavioral Science*, 21, 86-100.
- Pilyshyn, Z. Agnew, N. & Illingworth, J.: 1966. Comparison of individuals and pairs as participants in a mixed-motive game. In: *Journal of Conflict Resolution*, 10, 211-220.
- Rapoport, AN. & Chammah, M.A.: 1965, *Prisoner's Dilemma: a study of conflict and cooperation*. University of Michigan Press, Ann Arbor.
- Revenstorff, D., Wegscheider, R., Fitting, U. & Mai, N.: Markov models of gaming behavior in experimental non-zero-sum games. In: Kempf, W.F. & Repp, B.H.: *Mathematical models for the social psychology*. Huber, Bern.
- Schaefer, R.E.: 1976. *Probabilistische Informationsverarbeitung*. Huber, Bern.
- Schlenker, B.R. & Goldman, H.J.: 1978. Cooperators and Competitors in conflict. A test of the "Triangle Model". In: *Journal of Conflict Resolution*, 22, 393-410.
- Schulz, U. & Hesse, W.: 1978. Ein Modell für das Verhalten in Prisoner's-Dilemma-Spielen unter Berücksichtigung von Erwartungen. In: *Psychologische Beiträge*, 20, 551-570.
- Schulz, U. & Jonas, K.: 1982. Kooperation und Wettbewerb, Tit-for-Tat-Strategie und Kooperationserhöhung auf dem Hintergrund eines Modells für das Spielverhalten in wiederholten Prisoner's-Dilemma-Spielen.
- Schulz, U. & May, T.: 1982. Ein Modell für das Verhalten in Sequenzen von Konfliktspielen unter Berücksichtigung von Antizipationen und subjektiven Sicherheiten. Paper submitted to *Zeitschrift für Sozialpsychologie*.
- Schulz, U. & May, T.: 1983. Verhalten in Sequenzen von "Chicken"-Spielen unter Berücksichtigung von Antizipationen und subjektiven Sicherheiten. Paper presented on the 25. Tagung experimentell arbeitender Psychologen, Hamburg.

- Sermat, V.: 1967. The effect of an initial cooperative or competitive treatment upon a subject's response to conditional cooperation. In: *Behavioral Science*, 12, 301-313.
- Solomon, L.: 1960. The influence of some types of power relationships and game strategies upon the development of interpersonal trust. In: *Journal of Abnormal and Social Psychology*, 2, 223-230.
- Tedeschi, J.T., Steele, M.W., Gahagan, J.P. & Aranoff, D.: 1968. Intentions, prediction and patterns of strategy choice in a Prisoner's Dilemma Game. In: *Journal of Social Psychology*, 75, 199-207.
- Terhune, K.W.: 1968. Motives, situation, and interpersonal conflict within Prisoner's Dilemma. In: *Journal of Personality and Social Psychology*, 8, 1-24.
- Winkler, R.L.: 1967. The quantification of judgement: Some experimental results. *Proceedings of the American Statistical Association*, 386-395.

ON THE PROMINENCE STRUCTURE OF THE DECIMAL SYSTEM

WULF ALBERS

GISELA ALBERS

*Universität Bielefeld
Institut für Mathematische Wirtschaftsforschung
F.R. Germany*

Summary: This paper gives a first approach to a theory of the prominence of numbers of the decimal system.

Basic components of the theory are the limited rational principles of rule construction by iterated addition or subtraction of a given amount, the refinement of a scale by adding the means of any two neighbours, and its coarsening by omitting all uneven elements.

Applying these principles to the powers of 10 (i.e. the most prominent numbers of the decimal system) various scales with different degrees of exactness can be constructed. Empirical observations suggest that among different scales with about the same exactness generally that one is preferred which can be constructed in the easier way, where the construction principles are those from above and the coarsening procedure is assumed to have the same complexity as the two-fold refinement. This reduces the relevant scales to $(z \cdot a \cdot 10^n, z \in \mathbb{Z})$ with $n \in \mathbb{Z}$ and $a \in \{1, .5, .25 \text{ or } .2\}$.

The prominence of a number or a set of numbers is defined as the length of the steps of the coarsest scale containing the number or the set. (For empirical applications it is suggested to permit 25 percent exceptions.)

Empirical results indicate that the prominence of the set of numbers resulting from a specific decision situation is in many cases about one tenth of the numbers in question.

More precisely the prominence (or exactness of scale) of a specific decision problem seems to be selected according to the rule that it is maximal subject to the condition that the range of reasonable alternatives contains at least three prominent numbers.

Results of n-person bargaining experiments (with real valued characteristic function games) suggest that the evaluation of outcomes (more exactly, of pay-off differences) is done via the underlying prominence, such that the distance of two outcomes is evaluated by the number of prominence steps between them. By measuring incentives in this way a general theory of the probability that a "social" coalition (mainly an equal share coalition of all players) is formed could be developed for games of apex type, by which the empirical observation that the probabilities can be quite different if the same game is played on scales with different prominence can be explained. This result suggests that the prominence of the (spontaneously selected) scale can have basic importance for the analysis of individual or group decisions on numerical values.

1. *Prominent points of the decimal system*

The basic principles of constructing scales seem to be

- (a) the iterated addition (and subtraction) of a given amount x , so that all integer multiples of x are obtained
- (b) the refinement of a scale by adding the means of any two neighbours
- (c) the coarsening of a scale by omitting all uneven elements.

In the decimal system the main reference points are the powers 10^n ($n \in \mathbb{Z}$), and according to principle (a) each of these powers produces a scale ($z \cdot 10^n$, $z \in \mathbb{Z}$) which can be refined or coarsened.

The first refinement of such a scale according to (b) gives the numbers ($z \cdot 5 \cdot 10^n$, $z \in \mathbb{Z}$), the second produces ($z \cdot 25 \cdot 10^n$, $z \in \mathbb{Z}$). These types of refinement seem to be empirically relevant. - A third application of (b) gives ($z \cdot 125 \cdot 10^n$, $z \in \mathbb{Z}$), but we did not observe sets of empirical data (with the character of decimals, not multiples of $1/8$) in which the multiples of $.125$ were essentially more frequent than other multiples of $.025$. It seems that

the system $(z \cdot 1 \cdot 10^n, z \in \mathbb{Z}) = (z \cdot 10^{n-1}, z \in \mathbb{Z})$ is clearly preferred.

The coarsening principle (c) gives the systems $(z \cdot 2 \cdot 10^n, z \in \mathbb{Z}) = (z \cdot 2 \cdot 10^{n+1}, z \in \mathbb{Z})$ which can be chosen as an alternative of $(z \cdot 25 \cdot 10^{n+1}, z \in \mathbb{Z})$, since both systems have about the same degree of exactness. - It seems that in a fixed situation each individual selects one of these two scales. That both of them have some empirical evidence can be seen from the fact that some currencies have pieces of value .02, .20, 2, and 20 (as Switzerland), while others introduced .25, 2.5, and 25 instead (as the Netherlands). - The two-fold application of principle (c) gives $(z \cdot 4 \cdot 10^n, z \in \mathbb{Z})$ which, however, seems never to be used. The scale $(z \cdot 5 \cdot 10^n, z \in \mathbb{Z})$ with a comparable degree of exactness seems to be clearly preferred.

The result of these considerations are the scales

$$(z \cdot a \cdot 10^n, z \in \mathbb{Z})$$

where $n \in \mathbb{Z}$ and $a \in \{1, .5, .25 \text{ or } .2\}$.

The length of the step $\Delta = 10^n$ of such a scale is called its *prominence*. We call a real number x *prominent with* Δ if it is contained in such a scale with prominence $\Delta' \geq \Delta$. A set M of real numbers is *prominent with* Δ if all (in empirical applications 75 % of all) elements of M are prominent with Δ and Δ is maximal with this property.

The idea is that every decision of an individual for a number is made on one of these scales considering only numbers with a given prominence where the prominence seems to be selected sufficiently high such that the set of alternatives which have to be checked as possible solutions of the decision problem remains sufficiently low.

We mention the following deviations from the principle that only scales of the types above are selected.

(1) In situations with more than one decision maker it is frequently considered to split the joint outcome equally. This leads to additional prominent numbers which, however, have not been used to generate scales according to principle

(a) in the cases we observed.

(2) If the decision task is to divide a given amount x among different people, it can happen that prominent divisors of x are selected to generate scales using principle (a). (Some subjects selected a prominence of 30, when x was 150. - We explain this deviation from the theory by the assumption that the reference number x caused the subjects to leave the decimal system.)

(3) If the selected number is - as in price setting - such that others are supposed to react on them, then it can make sense to replace the numbers $(z \cdot a \cdot 10^n, z \in \mathbb{Z})$ by alternatives with a higher efficiency anticipating the reactions of the others. (Frequently 2.98, 2.78, 2.48, 2.28 are selected instead of 3.00, 2.75, 2.50, 2.25, such that the prices are kept just below the prominence steps of the buyers. But there are different rules of this type, also depending on the degree of reflection related with the act of buying, and we do not want to go into these details here. In the following we shall only consider examples with an obvious transformation rule.)

2. *The relative size of the prominence*

Empirical results give the impression that sets of data (as prices of comparable goods in the same shop, proposals in similar bargaining situations, etc.) quite frequently have a prominence of about 10 % of the numbers and the range is between 5 % and 25 %.

To give an impression of a set of possible prices resulting from such a behaviour, we consider as an example the case that each price is calculated with a relative prominence not below 5 %. (We suppose that each price is calculated separately and selected from a separate system $(z \cdot \Delta, z \in \mathbb{Z})$ with a prominence $\Delta = a \cdot 10^n$ which is as low as possible, but not below 5 % of the finally selected price.) Then the seller can choose among the following numbers in the range between 1 and $10 : ^1$)

¹) We assume that $a = .25$ is preferred to $a = .2$. - Possible prices between 10^n and 10^{n+1} are obtained by multiplication with 10^n .

1.0), 1.10, 1.20, 1.25, 1.30, 1.40, 1.50, 1.60, 1.70, 1.75, 1.80, 1.90 (A1)
 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75 (A2)
 5.00, 5.50, 6.00, 6.50, 7.00, 7.50, 8.00, 8.50, 9.00, 9.50 (A3)

In this system of possible prices there are structural breaks at 1.00 (where the prominence changes from .05 to .10, i.e. from 5 % to 10 %), at 2.00 (where the prominence changes from .10 to .25, i.e. from 5 % to 12.5 %), and at 5.00 (where the prominence changes from .25 to .50, i.e. from 5 % to 10 %). So the rule that no article is calculated with a higher relative prominence than 5 % causes relative prominences between 5 % and 12.5 %.

(Generally, a minimal relative prominence of α % causes relative prominences between α % and at most $2.5 \cdot \alpha$ % in the corresponding system of possible alternatives.)

Another example which is also comparable with observed price systems is the set of possible alternatives between 1 and 10, where each number is calculated with a minimal relative prominence of 10 %:

1.00, 1.25, 1.50, 1.75, 2.00, 2.25 (B1)
 2.50, 3.00, 3.50, 4.00, 4.50 (B2)
 5.00, 6.00, 7.00, 8.00, 9.00 (B3)

there the structural breaks are at 1.00, 2.50, and 5.00, where the relative prominences change from .10 to .25 (i.e. from 10 % to 25 %), from .25 to .50 (i.e. from 10 % to 20 %), and from .50 to 1.00 (i.e. from 10 % to 20 %).

These theoretical examples illustrate how prices with a relative prominence between 5 % and 25 % can occur. To underline the empirical evidence of these considerations we give some examples of observed price systems. Each of the examples gives all observed prices of a selected department of one shop or all prices of an advertisement or leaflet. (We remark that examples have been selected such that the prominence structure can easily be recognized.)¹⁾

¹⁾ All prices have been observed in shops at Bielefeld in the first months of 1983.

Example 1: slabs of chocolate (department store)¹⁾

ideal	.85	.90	1.00	1.10	1.20	1.25	1.30	1.40	1.50	1.60	1.70	1.75
real	.85	.89	.99 .98	1.08	1.18	1.23	1.28	1.38	1.48	1.58 1.59 1.60	-	-
frequ.	1	1	3	2	3	1	5	2	2	3	-	-

ideal	1.80	(1.85)	1.90	2.00	...	2.25(?)	...	2.75(?)
real	1.78, 1.80	1.85	1.90	1.98 2.00		2.30		2.70
frequ.	2	1	1	5		1		1

This price system fits quite well with the predictions of row (A1). Note that the structural break at 2.00 is also according to (A1).

It is interesting to note that price systems of different articles with different selling prices can have the same relative prominence. For example in two of the department stores we observed the prominence of television sets and portables was also comparable with (A1). They were prominent with 100 in a reange between 800 and 2000, where exclusively prices of the type $z \cdot 100 + x$ with $z \in \mathbb{Z}$ and $x = 99, 98$ or 48 were observed, and in one store $x = 48$ occurred only in 5 out of 64 cases.

Example 2: unpacked pralines (department store)

ideal + real	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50
frequency	1	1	4	3	1	3	3	5	-	8

ideal + real	4.75	5.00
frequency	7	1

This set is comparable with (A2). It should be noted that prices ending with .25 have been avoided and are essentially less frequent than those ending with .75 while in other cases the .25 values are preferred:

¹⁾ Articles offered with different tastes have been counted only once.

Example 3: leather blousons (department store)

ideal	200	225	250	275	300	325	350	375	400	425
real	195	-	250	-	295	325	350	-	395	425

Another candy has a considerably higher degree of prominence, comparable with (B3):

Example 4: chocolate eggs (department store)

ideal + real	40	(45)	50	(55)	60	(65)	70	75	80	...	100	...	120
frequency	10	1	4	2	14	1	7	11	3		1		1

That there are also price systems where the prominence 2 is preferred to 2.5 is illustrated by

Example 5: dress material (reduced prices of seasonal sales) (dept. store)

ideal + real	2	3	4	5	6	8	10	12	15	18	20	25	1)
frequency	28	24	7	91	17	76	19	18	11	6	4	1	

The relative prominence of these prices is comparable with (B) but it is exceptionally high at 2 and 3. This may be a consequence of the quick calculation in the sales situation or result from the principle to avoid broken numbers.

Another example of a price system according to (B) is

Example 6: shoes (discounter)

ideal	10.00	12.50	15.00	20.00	25.00	30.00	35.00	40.00	50.00	60.00
real	9.99	12.98	15.98	19.98	25.98	29.98	35.95	39.95	49.95	59.95
					25.00		35.00		49.50	59.50
					24.98					
ideal	70.00	75.00	80.00	90.00	100.00	110.00				
real	69.95	75.95	79.95	89.50	-	109.00				

The last example gives a price system which is somewhat more exact than (B). It permits insight into some additional conditions under which numbers with a lower prominence are selected:

Example 7: all articles of a leaflet with do-it-yourself materials and equipment (discounter) ¹⁾

ideal	1.25	2.00	2.50	3.00	3.50	4.00	5.00	5.50	6.00	(6.25)
real	1.29	1.99	2.49	2.99	3.49	3.99	4.99	5.49	5.99	6.29
frequ.	1	1	1	3	1	1	5	1	3	1

ideal	6.50	7.00	7.50	8.00	9.00	10.00	12.00	12.50(?)	15.00
real	6.49	6.99	7.49	7.99	8.99	9.99	11.98	12.98	14.98
frequ.	1	3	3	2	7	5	1	3	2

ideal	17.00	17.50(?)	19.00	20.00	22.00	25.00	27.00	30.00	35.00
real	16.98	17.98	18.98	19.98	21.98	24.98	26.98	29.98	34.95
frequ.	1	2	2	7	2	4	1	3	3

ideal	40.00	45.00	50.00	55.00	60.00	65.00	70.00	75.00	80.00
real	39.95	44.95	49.95	55.00	59.95 ²⁾	65.00	69.95	75.00	79.95
frequ.	3	4	4	1	5	1	5	1	2

ideal	85.00	90.00	100.00	110.00	120.00	130.00	140.00	150.00	160.00
real	85.00	89.50	99.50	-	119.00	129.00	139.00	149.00	-
frequ.	1	1	4	-	2	1	3	4	-

ideal	170.00	180.00	190.00	200.00	... 240.00	... 260.00
real	169.00	179.00	189.00	198.00	239.00	259.00
frequ.	3	1	1	2	1	1

It should be remarked that the prices with a comparably high degree of exactness are set for articles to which similar or comparable articles existed with

¹⁾ 15 articles from the food section which were in a separate board have been omitted.

²⁾ 1 article had the price 59.50.

prices with a higher prominence: curtain rails (with different lengths) had the prices 18.98 (120 cm), 19.98 (140 cm), 21.98 (160 cm) and 24.98 (180 cm), where the 140 cm and the 180 cm versions give the reference points 20.00 and 25.00. The prices of spotlights were

	pine	oak
single	19.98	21.98
two on a beam	39.95	39.95
three on a beam	55.00	59.95
three on a rondell	75.00	85.00

It can be easily realized that the price of the single pine version sets the reference point 20.00. The single oak version is calculated one step (10 % = 2.00) higher. The prices of the double and threefold versions are set cheaper than the corresponding multiples of the single versions, however, the selected degree of prominence did not permit a reduction of the double pine version. The price of the pine rondell version is again a number with a high prominence (75), whereas the oak rondell got a price which is one step (10) higher than 75.

These two examples explain most of the prices with a comparably low prominence, namely 18.98, 21.98 (2x), 55.00 and 85.00. Also the price of 6.29 (per **square** meter of rock wool) results from a price of 5.00 per running meter for the normal width of .80 meters.

3. *The prominence selection rule*

The observation that goods with the same sales price but different margins differ in the exactness of calculation suggests another approach to the prediction of the prominence of price or value systems. The main criterion on this approach is the range of reasonable alternatives of the decision problem.

prominence selection rule: For a given range [x, y] of reasonable alternatives the prominence is selected in such a way that the range includes at least three numbers with his prominence.

Although this rule seems quite tough it points better predictions than the

10 % - rule of the preceding paragraph. To explain the behavioral ideas behind this rule we consider an example of price setting :

A seller reflects upon the adequate sales price of an article. From his experience he has some idea of possible prices on the market. He can describe the corresponding range of prices roughly by an upper and a lower bound. Generally, these bounds have a comparably high prominence. Now our assumption is that the seller selects the prominence of the scale in such a way that besides the extremes at least one alternative between them becomes prominent.

Figure 1 gives range and prominence for a selection of articles from the food market. The corresponding data were collected in 68 shops at Bielefeld in May 1982. We selected all branded articles of this survey which had a selling price above 2.00 and for which the prominence structure was not perturbed by a price recommendation. The range of reasonable alternatives is estimated by the range of observed prices omitting the 10 % tails of both sides.

In figure 1 the closed points refer to branded articles (special brands of coffee, honey, brandy, champagne, instant food and pre-cooked meal) while the circles belong to other goods (apples ("golden delicious"), tomatoes, cheap liquor and sliced bread of a certain quality).

The results show that the range has a size between 2Δ and 5Δ where Δ is the prominence of the set of actually selected prices. This supports the prominence selection rule permitting a maximal extension of the range of reasonable alternatives between 2Δ and 5Δ , if the boundaries of the range are assumed to have at least prominence Δ . To illustrate this: if the range of reasonable alternatives is $\{x, 2 \leq x \leq 4\}$ then $\Delta = 1$ and the range has an extension of 2Δ , if the reasonable alternatives are $\{x, 1.5 \leq x \leq 3.5\}$ then $\Delta = .5$ and the range has an extension of 5Δ .

It is interesting that the nonbounded articles have a larger relative range ($> 5 \Delta$, the difference to the branded articles is significant on the 0.02 level). This can be explained by the fact that these articles do not have a comparable homogeneity of calculation. The only exception (cheap coffee) which is unbranded but has a relative range of 4.3Δ seems to have a quite homogeneous calculation.

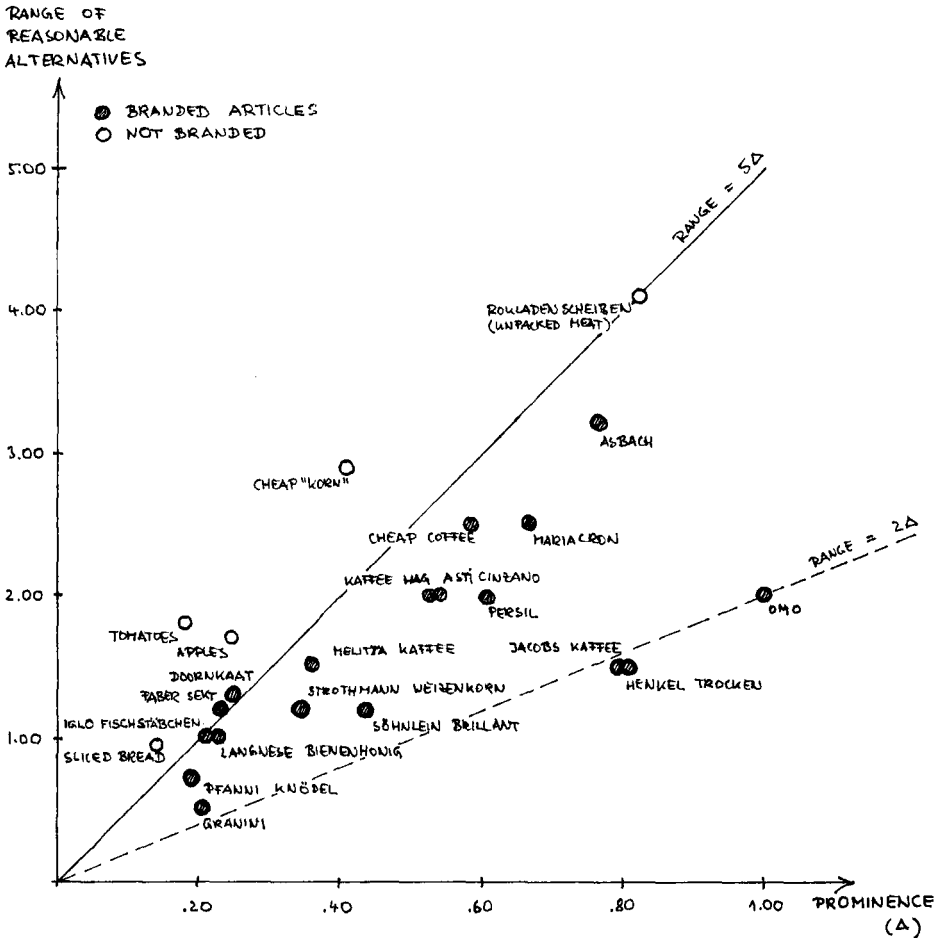


Figure 1: Range of reasonable alternatives versus prominence for some branded and unbranded articles

4. Evaluation of payoff differences by means of the prominence

Results from game theoretical experiments indicate that the prominence structure of a decision problem does not only serve to select possible numerical alternatives for a decision but is also used to evaluate differences between possible outcomes. The following results which shall be reported in short seem to admit this interpretation.

We refer to a series of n -person characteristic function games with free communication and transferable payoffs. Each of these n -person games was played by n^2 players. Each player played the game n times in n subsequent rounds meeting no other player twice.

The games can be described by the set of players $N = \{1, 2, \dots, n\}$ (here $n = 4$ or 5) and the payoffs of the coalitions $S \subseteq N$ which can be distributed among the players of the coalition. Here these payoffs were of the following type:

$$v(S) = \begin{cases} A & \text{if } 1 \in S \text{ and } S \setminus \{1\} \neq \emptyset \\ B & \text{if } S = N \setminus \{1\} \\ 0 & \text{otherwise} \end{cases}$$

where $A \geq B \geq A/n$ are positive real numbers.

In these games player 1 has a relatively strong position since he needs only one other player as a partner while the other players can coalesce without player 1 only if all of them stick together. However, this symmetric weakness of players 2, 3, ..., n causes them to consider the coalition $\{2, 3, \dots, n\}$ as their first coalition alternative which is strongly supported by social arguments (compare ALBERS [1978]). Player 1 can try to corrupt one of the other to form a coalition with him.

Within the "social" coalition $\{2, 3, \dots, n\}$ the payoff was distributed equally. In many cases this coalition has been extended to the grand coalition $\{1, 2, \dots, n\}$ again according to the equal share principle. There are two arguments for such an extension: one is not to exclude player 1 for social reasons, and the other is that the coalition is safer if no player is outside (the grand coalition N has not been broken in any of our experiments).

A central problem of the game was whether a social coalition should be formed or one of the two-person coalitions. The incentives for this decision can be characterized ex post by the mean payoffs x_1, x_i of player 1 and his partner in the two-player coalition and the outcomes a_1, a_i ($i = 2, 3, \dots, n$) in the social coalition. Here we set $a_1 = a_i = A/n$ since in nearly all cases where a social coalition became

stable the grand coalition N was formed. (Only in one sequence of games the players preferred $\{2, 3, \dots, n\}$ to the grand coalition. For this sequence we defined a_i by $B/n - 1$.)

It seems reasonable to assume that the incentive of a player j to leave the social coalition and form a two person coalition is proportional to the corresponding payoff difference $x_j - a_j$. Empirical results show that this incentive (which can be measured indirectly by the proportion of non-social results in the final outcomes of the games) remains more or less unchanged if the constants A and B defining the value structure of the game are simultaneously replaced by $10A$ and $10B$. On the other hand, they could be essentially influenced by the degree of involvement of the players in the game (measured by the relative prominence of the set of numerical proposals).

Both results can be explained by the assumption that the evaluation of the payoff differences is influenced by the prominence in such a way that the incentives to leave the social coalition are

$$d_j = \frac{x_j - a_j}{\Delta}$$

where Δ is the prominence of the analysis (measured by the prominence of the set of proposals). Of course, Δ should be measured individually but we had not enough observations for such an analysis. We, therefore, assumed that all players analyse the game with more or less the same prominence and we estimated Δ for each type of the game by the prominence of the set of all proposals of all players.

By this formula the incentive is given by the number of prominence steps between x_j and a_j , for instance if $x_j = 500$, $a_j = 300$ and $\Delta = 50$, then the incentive is 4, since there are 4 prominence steps (350, 400, 450, 500) from 300 to 500.

If in the same situation with the same expected outcomes another player (or group of players) only plays with a prominence $\Delta = 100$, then the incentive to leave the social coalition is only half as high, since for this

prominence there are only two steps (400 , 500) from 300 to 500.

It can be seen from table 3 that in two cases the same game has been played with different prominences. In both cases the corresponding motivational situations became completely different and very different degrees of cooperation resulted.

We assume that the probability p_j of a player j to decide for the two-person coalition is proportional to his incentive d_j and our data show that the proportionality factor has the size of about .1 (least square analysis gives an optimal value of .094). So we obtain

$$p_j = .1 \cdot d_j = .1 \cdot \frac{x_j - a_j}{\Delta}$$

(since p_j is a probability, it must be replaced by 1 if the formula gives a higher value.)

Assuming that the decisions of the players between a two-person coalition and a social coalition are made independently, we get the following formula for the probability that a nonsocial coalition is formed.

$$p = p_1 \left(1 - \prod_{i=2}^n (1 - p_i) \right)$$

where $\prod (1 - p_i)$ gives the probability that all weak players decide to form a social coalition, $1 - \prod (1 - p_i)$ is the probability that at least one weak player decides for a nonsocial coalition and the final expression selects those cases where player 1 also decides to play in a nonsocial way.

Figure 2 gives the results of 8 such series of experiments. (They involved 155 subjects who played 155 games.) Δ has been measured as the prominence of the respective sets of proposals, x_j as the mean outcome in the two-person coalition of player 1 with another player, a_j is the equal share in the social coalition. Using these values our theory gives

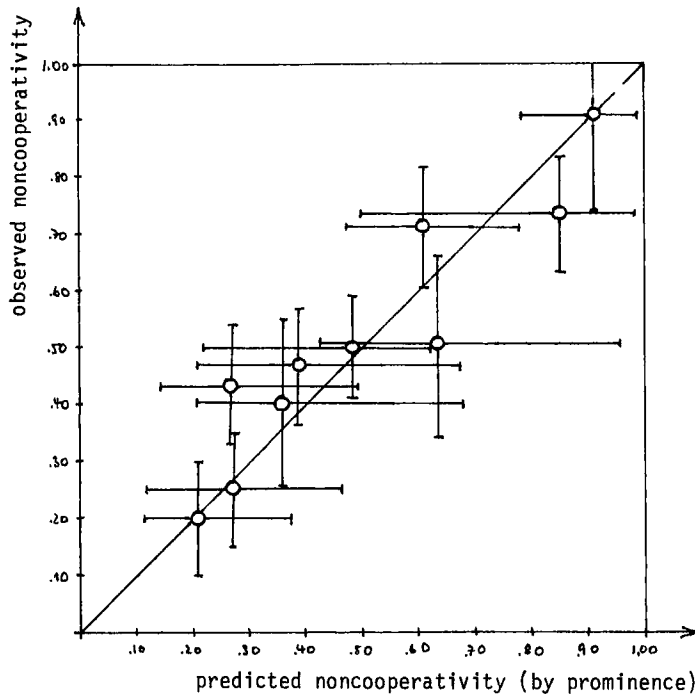


Figure 2: Predicted (by prominence) and observed noncooperativity in games of Apex type

a prediction of the proportion p of noncooperative results. On the other hand, p can be measured by the percentage of nonsocial results if outcomes which do not involve the social coalition or the two-person coalition (these were less than 10 % of the results) are omitted.

The error of the percentage of nonsocial results is estimated by the assumption that 1.5 of the 15 - 25 results of a sequence of games could have a social instead of a nonsocial result or vice versa. The error of the prominence is estimated by the probability according to the geometric mean of the observed prominence and the next higher (respectively lower) promi-

n	v(N)	v(234)	prominence	noncoop. results	
				predicted	observed
5	100	100	5	.47	.40
5	1000	1000	20	.64	.71
5	1000	1000	50	.92	.95
5	1000	1000	50	.70	.50
5	1000	1000	100	.24	.20
4	100 ¹⁾	75	10	.51	.50
4	160	165	5	.87	.73
4	160	165	10	.31	.43
4	320	315	20	.32	.25
4	1000	1000	50	.44	.47

Table 3: Predicted (via temptation) and observed degree of noncooperativity in games of Apex type.

nence. The error of the mean outcome x_j could be neglected because it was comparably small.

The data give a correlation of .83 of observed and predicted outcomes which is significant on the .01 level.

Maybe it cannot be finally decided whether the prominence really influences the evaluation of the payoff differences $x_j - a_j$ or if both, the evaluation of the differences and the prominence, depend on an underlying behavioral pattern. However, the observed results indicate that the prominence influences the evaluation since the error of p caused by the inexactness of Δ seems to be essentially lower than admitted by the "natural" error bounds given by the geometric means to the next higher or lower prominences.

¹⁾ In this game we set $v(1) = 25$

References

- Albers, W.: 1983. Bloc Forming Tendencies as Characteristics of the Bargaining Behavior in Different Versions of Apex Games, in H. Sauermann (ed.): *Coalition Forming Behavior*, Tübingen, 172-203.
- Albers, W. and Albers G.: 1983. Prominence and Aspiration Adjustment in Location Games, in R. Tietz (ed.): *Aspiration Levels in Bargaining and Economics Decision Making*, Berlin, Heidelberg, New York, 243-258.
- Albers, W. and Albers, G.: 1983a. Selection of optimal numbers of the decimal system in individual decision making, Working Paper of the Institute of Mathematical Economics, University of Bielefeld.
- Emory, F.: 1969. Some Psychological Aspects of Price, in Taylor-Wills (eds.), *Pricing Strategy*, Princeton.
- Ginzberg, E.: 1936. Customary Prices, *American Economic Review* 25, 296.
- Gabor, A., Grange, J.W.J., Sowter, A.P.: 1971. Comments on "Psychophysics of Prices, *Journal Marketing Research* 8 (May), 251-252.
- Crößmann, H.J., Tietz, R.: 1983. Market Behavior on Aspiration Levels, in R. Tietz: *Aspiration Levels in Bargaining and Economic Decision Making*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 170-185.
- Kamen, J.M., Toman, R.J.: 1970. Psychophysics of Prices, *Journal of Marketing Research* 7 (February), 27-31.
- Kamen, J.M., Toman, R.J.: 1971. "Psychophysics of Prices" :A Reaffirmation, *Journal of Marketing Research* 8 (May), 252-257.
- Monroe, K.B., Venkatesau, M. : 1969. The concept of price limits and psychophysical measurement: A Laboratory Experiment, in *Proceedings Fall Conference*, American Marketing Association, Chicago 1969.
- Oesterwalbesloh, K.H., Rembde, M. (in preparation): Prominenzstrukturen und Preissetzungsverhalten im Einzelhandel, diploma thesis, Institute of Mathematical Economics, University of Bielefeld
- Oxenfeldt, A.R.: 1968. How Housewives Form Price Impressions, *Journal of Advertising Research* 8 (September), 9-17.
- Sauermann, H., Selten, R.: 1962. Anspruchsanpassungstheorie der Unternehmung, *Zeitschrift für die gesamte Staatswissenschaft*, 118, 577-597.
- Schelling, T.C.: 1960. *The Strategy of Conflict*, Cambridge.
- Schmitz, G.: 1964. Zwischenbetrieblicher Vergleich der Einzelhandelspreise sortengleicher Konsumwaren, Köln, Opladen.
- Simon, H.A.: 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69, 99-118.
- Tietz, R.: 1978: Entscheidungsprinzipien der bilateralen Anspruchsanpassung, in: E. Helmstädter (ed.): *Neuere Entwicklungen in den Wirtschaftswissenschaften*, Schriften des Vereins für Socialpolitik, Neue Folge, Bd. 98, Berlin, 431-455.

This Page Intentionally Left Blank

DEVELOPMENT AND EPISTEMOLOGY

This Page Intentionally Left Blank

CONCEPTUAL AND THEORETICAL ISSUES IN DEVELOPMENTAL RESEARCH ON THE ACQUISITION OF THE PROBABILITY CONCEPT¹⁾

ROLAND W. SCHOLZ

MANFRED WALLER

*Universität Bielefeld
Institut für Didaktik
der Mathematik
F.R. Germany*

*Universität Bielefeld
Abteilung für Experimentelle
und Angewandte Psychologie
F.R. Germany*

Psychological theories are being developed about the cognitions or attitudes etc. towards certain objects, or about behavior in certain situations. When developing a psychological theory one should clarify "for what", i.e., about which object or for which situation it is intended. This is also true for a theory about the acquisition of the probability concept. However, one has to realize that this has not been done - with the exception of PIAGET and INHELDER (1951/1975). As the probability concept can be interpreted in essentially different ways we begin with an outline of the (main-)variants of the probability concept in section 1. Before turning to an analysis of the developmental theoretical approaches into the acquisition of the probability concept, we pick out which types of probability are investigated in section 2.

Before the theoretical research approaches on the ontogenesis of the probability concept can be analysed, we have to start with a slightly more detailed study of the probability concept and how it is interpreted within developmental psychological research. This seems to us to be necessary, as a comprehension of a concept should precede an understanding of the ontogenetic development of this concept.

1) We wish to thank Prof. Ian Hacking for his helpful comments and for his and Jonathan Harrow's assistance in producing a readable English text.

1. *Variants of probability*

The main dividing line in interpreting probability is still that between the *objectivist* and the *subjectivist* interpretations of this concept. These can be identified since the *emergence of probability*. "Probability has two aspects. It is connected with the degree of belief warranted by evidence, and it is connected with the tendency displayed by some random devices, to produce stable frequencies." (HACKING, 1975, p.1)

Both sides of probability are, in our opinion, indispensable and must be integrated into an understanding of probability that encompasses both aspects. Nevertheless, these different interpretations are frequently considered to be complementary by the schools of thought disseminating them.

A prototype of the *objectivist* conception is the so-called frequentist approach of Richard von MISES (1928). For this position, it is fundamental to comprehend probability as the limit of a relative frequency established by an experiment which can be repeated at will. Beyond that, it is decisive for this interpretation that these frequencies are observable and independent of each other, and that the mechanism of chance is explicit.

From a puristically objectivistic point of view, combinatorical aspects are inessential. They merely can play a part in model considerations. Urns, dice, and coins as idealized symmetrical objects are the models for certain random situations fulfilling the objectivist criteria of probability. Hence such calculations as those playing a central role in LAPLACE's classical concept of probability are only admissible if the elementary probabilities prove to be equally probable in the frequentist sense.

For the *subjectivist*, the following statement holds true: "Subjective probability is one's degree of belief in an outcome based on an evaluation making the best use of all the information available to him and his own skill." (DE FINETTI, 1974, p. 16) The basic idea establishing the subjectivist theory of probability is the betting or odds paradigm which relates

the probability concept to the concept of utility. However, one has to regard the theory of personalist or subjectivist probability as dealing with *idealized* persons. These persons are consistent and free of logical contradictions in their judgments. It is admissible, however, to construe different idealized persons who differ in their opinions about one and the same situation. Contrarily, within CARNAP's logical theory of probability for a given situation only one rational person can be constructed relevant to a given body of information.

The subjectivist conception, however, also provides a definite interpretation of classical probability experiments like urns or tossing coins, if the information is exhaustive and the situation is definitely interpreted by means of symmetry. Under this condition, the subjectivist theory deals with one person. "In those situations where the subjectivist believes there is symmetry he would act as if he accepted classical probability, for example leading to the Laplacian ..." (FINE, 1973, p. 234f)

We should like to point out here the *difference between mathematical probabilities* and the *probabilities which psychologists attempt to measure* by different methods and means in *real* individuals, in experiments or under field conditions. Mathematical theory must be considered as a logic of decisions, while the psychological access to "subjective" probabilities *must be considered* as an attempt to determine an individual's cognitions and judgmental processes in which mathematical concepts, especially mathematical probabilities, may be included.

It is by no means expected that a real individual is free of contradictions and consistent in his/her judgments, even if he/she uses certain elements of a mathematical theory. It must be noted that the "subjective probabilities" studied in psychology may be made the object of a contradiction-free, consistent mathematical theory.

As psychologists we believe that a human being as an individual is actively reconstructing his environment with the help of his information and knowledge, thus assuming that "normal people" build a model of their environment in the same way as scientists.

Within the subject's model, subjective probability may also be assigned to events which in the objectivist sense are not probabilistic. "The adjective 'subjective' refers", here, "to the subject and not to the arbitrary assessment which is independent of the objective measure. Subjective probabilities in psychology ... are to be understood as the objective measure of subjective assessments." (HÖRZ, 1980, p. 190, translation by the authors) These subjective assessments according to HÖRZ (1980) are the result of a 'Widerspiegelung' (process of reflexion). "Within this process, verified theories, empirical facts, models of various kinds, and hypotheses play a part." (HÖRZ, 1980, p. 190)

2. The probability concept within developmental research

After these fundamental considerations we will now turn to the question how these interpretations of the probability concept as distinguished above are dealt with in the investigations into the acquisition of the probability concept. We should like to start this section with the following thesis:

In research into the acquisition of the probability concept, only the objectivist concept of probability has (with one exception) been studied.

Now for the justification of this thesis. Traditionally, one can distinguish between two paradigms: research into the *understanding of classical chance experiments*, and research into *probability learning*.

Developmental research into the acquisition of the probability concept was initiated by PIAGET and INHELDER (1951, see also PIAGET, 1950). In their experiments they predominantly used random situations typical for the objectivist probability concept. Most frequently one finds urn or container experiments and spinner tasks. The situations are prototypes of tasks where in the probability of events is determined either spatio-temporally, or logico-arithmetically. In the case of the container experiment, there is the classical distinction between the so-called one container and the two container task. One distinguishes between single or multiple drawing with

and without replacement. Accordingly, experiments of the spatio-temporal type were carried out with the "one spinner" task on the one hand, and with the "two spinner" task on the other.

As a whole, the repertoire of chance situations used in developmental experiments is rather limited. Hardly any experimental task other than the classical one has been designed. It is particularly bad that there are no attempts to use situations with explicit chance mechanisms without symmetry.

Let us now turn to the second paradigm, to probability learning. A particularly revealing description of probability learning which shows the amazingly "broad" scope of this branch of research is given by MESSICK and SOLLEY (1957, submissions by the authors). "The systematic change in output probability as a function of having more and more experience with an input is called probability learning. Studies in this area have used either adult college students ... or subhuman species such as rats ... or goldfish ..." What this definition shows is that the context-free task interpretation was dominant for a long time. The subjects were confronted with devices (black boxes) and were asked to predict the flashing on of lamps etc. without being informed about the inside of the apparatus and thus about the nature of the task. The implications of this task interpretation were pointed out rather early by FLOOD (1954). Even at the beginning of research into probability learning it was noted in fact that subjects, in tasks with two alternative stimuli, do not always choose the more frequent alternative. Instead of this, they adjust their response frequencies to the frequencies of the stimuli. This response behavior is denoted as probability matching. Comparing subject performance by the optimal prediction strategy in stationary random sequences it was evaluated as being sub-optimal or fallacious. With regard to this evaluation, we should like to quote two points revealing the main problems associated with research into probability learning.

- 1) Tasks are cognitive demands and thus co-determined by an individual's active reconstruction. One and the same situations may be perceived very

differently. As GOODNOW (1955) was able to show, the deviation of an individual's performance from the optimal response strategy may often be explained by assuming that subjects do not recognize random sequences as such, but instead try to discover and predict regularities and contingencies in the sequence by means of a problem-solving strategy (cf. WEIR, 1964).

2. Individuals tackling probability learning tasks and hypothesizing that these are "genuine random sequences", generally are not so wrong in their assumption. This is because in the experimental design of probability learning tasks not all sequences are usually permitted, but only those which "make a chance impression".

Under certain conditions, this may have the consequence that it is sometimes better not to select the more frequent alternative in 100 percent of the cases.

Now let us return to our thesis. The sequences of stimuli in probability learning tasks meet important criteria of random sequences in the objectivist sense. This is especially true with regard to probability learning tasks in developmental research. To infer from this that the subjects perceive the tasks in the same objectivist way is inadmissible. As we have already pointed out, the subjects in the experimental situations were often not informed about the nature of the task and about the way events were being generated. Hence, deterministic reconstructions of the tasks were not unreasonable. This position is taken up for instance by WEIR (1964).

We mentioned above that there is only one work in this field that refers to the subjectivist concept of probability. As reported by Lee ROSS (1980), one of his students (JOSEPHSON, 1977) presented children's descriptions of persons acting in the "Boy meets Dog" story. First, children received a description of a person of the following kind:

"This is a picture of Johnny. (E shows S a photograph of a child.) When

Johnny goes to the park with his mother, he doesn't go on the high slippery slide, because he's scared. When he's at school, he doesn't play with kids who are bigger than he is, because sometimes they push him around. Johnny is also scared of the dark." (JOSEPHSON,1977, quoted according to ROSS, 1981, p. 30)

After that, children got a description of a situation:

"Now, here's a picture of a dog. (E shows S a photograph of a large German Shepherd.) Suppose Johnny was playing by himself in front of his house one day, and this dog in the picture came running up to him and barked and growled pretty loudly." (JOSEPHSON,1977, quoting from ROSS, 1981, p. 30)

There were four versions of the story. After each version, the children were asked to give a judgment about the expected reaction of the actor, and "then to indicate how sure they were about their answer." (ROSS,1981, p.21) It seems obvious that this type of task is something quite different from container experiments and probability learning tasks. Stories like "Boy meets Dog" have nothing more to do with the probability concept, in the objectivist sense because no chance mechanism at all can be determined.

3. Developmental theories and research approaches

3.1 The cognitive-developmental approach of PIAGET and INHELDER

The most comprehensive research approach dealing with the development of the probability concept is without doubt to be found within the cognitive-developmental theory of Jean PIAGET (1950). More than thirty years ago, PIAGET and INHELDER (1951) systematically analysed the probability concept in children and formulated a theory to explain it's development.

It is well known that, according to PIAGET, cognitive development consists of a sequential transformation process during which elementary sensori-motor schemata are restructured into progressively more complex cognitive structures, initially into pre-operative and then into operative ones.

Within PIAGET's theory operative cognitive structures constitute the most important developmental prerequisite for the emergence of logical-mathematical concepts. This is because the growing child is only by means of logical operations able to discover and to reconstruct the quantitative-logical relations inherent in problems. This also applies to the child's construction of the probability concept. According to PIAGET and INHELDER the very basis for the acquisition of this concept is the ability to distinguish between chance and necessity.

The pre-operative child is unable to make this distinction. At this stage of development the child still lacks the ability to construct logical relations such as cause and effect or similar deductive relations, that are necessary for an understanding of why events occur. In other words, the pre-operative child has not yet developed the cognitive frame on the basis of which he/she can grasp the difference between necessary (predictable) and random (non-predictable) events. Random events, in his/her understanding are still subject to the same deterministic order, which also controls necessary events. In order to be able to understand the necessity versus the randomness of events, the child must have organized cognitive structures into concrete operations of thinking. These enable the child to comprehend the contingencies of deterministic events by means of building logical relations, and thus to recognize the logical necessity of their occurrence. On the other hand, the contingencies of stochastic events cannot be represented in concrete-operative structures because of their irreversibility. Hence, the child cannot infer their occurrence, and by recognizing the distinct predictability of necessary versus random events he/she becomes able to differentiate between them.

In a further developmental step beyond this conceptual distinction, the child acquires the ability to make chance events operatively calculable. The concrete-operative child may already be able to make some quantitative estimates of probability, but only, when the number of events is limited, and when, simultaneously, the proportion between favourable and non-favourable cases can be constructed by additively linking simple part-whole-relationships. In order to overcome these limitations that are set upon con-

crete operations in probability calculus, the child has to be able to construct and represent the totality of all events and to take multiple relations into account when he/she is calculating proportions. The pre-requisites for doing this are acquired in the formal-operative stage, during which the concrete operations are co-ordinated and transformed into a completely closed and reversible system of formal-logical operations.

3.2 FISCHBEIN's learning developmental approach

According to FISCHBEIN (1975, 1976), PIAGET and INHELDER's stage model of the development of the probability concept has two serious deficiencies: First, the findings on the growth of probability learning (cf. GOULET & GOODWIN, 1970), particularly those made in the sixties, cannot be integrated into this model. Secondly, in FISCHBEIN's opinion, PIAGET and INHELDER underestimate the importance of the learning processes leading to the formation of the probability concept. As an alternative, FISCHBEIN formulates a learning-developmental approach to the acquisition of the probability concept. This approach is based on studies on the growth of probability learning that reveal that children already show probability matching at preschool age. Taking these findings into account, FISCHBEIN postulates that children, even in the pre-operative stage, have a pre-conceptual understanding of both relative frequencies and probabilities at their disposal. This pre-conceptual understanding is founded on *primary intuitions*. In FISCHBEIN's opinion primary intuitions are a type of implicit knowledge that grows spontaneously out of a child's everyday experience, for example from the perception of stochastic events in everyday life situations. As the child grows older primary intuitions that, according to FISCHBEIN, are developed with regard to the concepts of *relative frequency*, *chance* and *probability* are successively transformed into an operative concept of probability. This transformation process is not the spontaneous and quasi-automatic result of a self-regulating process of growth, but rather is mediated by instructional intervention at school (for instance by explaining to the child the basic ideas of probability theory, by teaching strategies and rules of probability calculus, etc.). Without such purposeful intervention, an operative concept of probability cannot be developed, even in adults.

3.3 Information processing approaches

Within the frame of information processing theories, three recent approaches can be identified that are independent developments based on heterogeneous theories of cognitive development. They provide a much more radical alternative to the cognitive-developmental stage model than FISCHBEIN's learning-developmental approach. These approaches are all built on a proposal that is contrary to PIAGET and INHELDER. According to this proposal the acquisition of probability competencies is not a function of the development of cognitive structures resp. their organization into operative systems, but rather depends, on both the growth of information processing space, and on the changing of information processing and problem solving strategies that occurs during the course of development.

Within the frame of the information processing approaches, a comprehensive and unitary theory of the acquisition of the probability concept has not been formulated until now. There are, however, some separate studies that focus on the developmental analysis of strategies and the abilities in solving probability problems.

SCARDAMALIA (1977) attempted to show that the ability to cope with combinatorial problems is a function of the interaction between a subject's available information processing space and the information processing demands of a respective problem task. Subjects with different information processing spaces were asked to work on combinatorial problems which, according to the results of a task analysis, were designed in such a way that, holding their logical structure invariant, their information processing demands varied. Results showed that even concrete-operative children are ready to solve combinatorial problems provided that the task demands do not exceed their information processing space. Contrary to PIAGET and INHELDER's supposition, SCARDAMALIA's findings indicate that formal-operative thinking is not a prerequisite for being able to cope with combinatorial problems.

BRAINERD (1981) proposed that the development of probability judgments

in container tasks is adequately explained as being a function of the growth of working memory. Taking into account the available evidence on the mnemonic representation of frequency data reported by BOGARTZ (1969), BRAINERD at first designed a general model of the information processing taking place in working memory during the solving of container tasks. This model contains a total of four storage and three processing operations.

In a second step, he modified this model with regard to several variations of the container task (for instance with regard to tasks with-versus-without replacement), in such a way that specific effects of each of the four storage and the three processing operations on subject's responses could be predicted and tested by analysing the frequency of hits and failures observed. Starting from these predictions, BRAINERD conducted a series of twelve experiments with six variations of the container tasks. He used pre-school and primary school children. Children's responses indicated that all age groups did well in storing frequency data. However, younger children failed to retrieve this correctly stored data. In other words, the wrong predictions observed were primarily due to retrieval errors. As BRAINERD proposes, these retrieval errors are not dependent on a lack of specific cognitive structures, but rather on limitations of memory space. Therefore, in his opinion the decrease of failures in older children is sufficiently explained as a function of their growing memory space.

Finally, proceeding from his *rule-assessment-approach*, SIEGLER (1981) attempted to prove that information processing rules for solving probability tasks such as card deck experiments change in the course of a child's development. SIEGLER postulates that children apply four rules with an increasing degree of complexity, both when working on probability tasks in the card deck experiment and when solving non-probabilistic problems. These rules are defined in the following way.

- Rule I: Only the frequencies of the favourable elements are observed. Accordingly, the deck containing the larger number of favourable elements is chosen.
- Rule II: Rule I is adhered to except when the frequency of the preferred elements is the same for both decks. In this case, the frequency of the non-favourable elements is also taken into account.

Rule III: The difference between favourable and non-favourable elements is determined for each deck, and the deck with the larger algebraic number is chosen.

Rule IV: The proportions of favourable versus non-favourable elements in both decks are calculated.

According to SIEGLER the growing child acquires these four rules step by step. As the cognitive development is proceeding they are integrated into increasingly complex systems of rules analogous to binary decision-tree models. Which rule is applied in solving a probability problem will depend both on the complexity of a subject's system of rules, and on the structure of the task to be solved (e.g., on the relative frequencies of favourable versus non-favourable elements).

In order to test the validity of his assumptions, SIEGLER designed a total of six deck tasks. Favourable versus non-favourable elements were combined in such a manner that the percentage of hits for any of the four rules could be predicted within all tasks with respect to each of the four rules. These predictions were then compared with the task responses of 3- to 20-year-old subjects.

Subject's responses agreed very well with the predictions made for Rules I and IV. However, they only partly fitted the response patterns that were expected for Rules III and II. Furthermore, it was quite obvious that pre-school children, with one exception, were applying only Rule I, whereas even the 8-year-olds, and more pronouncedly the older subjects, made their probability judgments predominantly according to Rule IV and partly to Rule III, thus quasi skipping Rule II.

4. Discussion

In a study of the theories and research approaches outlined above one can find only slight disagreement as to the age period within which children should be ready to work successfully with probability tasks. However, there is much divergence of opinion as to how children cope with such tasks.

The main issue has been what weight should be assigned to conceptual knowledge versus strategic abilities with regard to the solving of probability problems. PIAGET and INHELDER definitely presume that only by acquiring the probability concept in the course of the development of operative-cognitive structures, will the subject be enabled to solve the probability tasks in an adequate way, i.e. by means of logical and combinatorial thinking. The validity of this assumption, however, seems to be questionable. At first, it is disputed by the above mentioned findings of SCARDAMALIA (1977) which clearly indicate that correct solutions of combinatorial problems do not necessitate formal operative thinking. Secondly, ROSS and HOEMANN (1975) were able to show that formal operative adolescents only rarely apply combinatorial strategies when working on probability tasks in a container-related experiment, but still often use strategies which characterize the thinking of the concrete-operative child. A similar problem-solving behavior may be shown by adults in proportional reasoning within everyday contexts (cf. CAPON & KUHN, 1979). These and other findings strongly suggest that even when competencies are available for formally correct solutions, these competencies may not be realized immediately, due to missing solution strategies. Hence, they very strongly emphasize the role of strategic abilities in solving probabilistic problems, and therefore they point to the neglect of just these abilities by PIAGET and INHELDER. Additionally, a lot of evidence seems to indicate that the development of the probability concept may be enhanced by the asking of children to work on probabilistic problems and also by the strategical knowledge which they acquire during problem solving.

Taken in combination, the research findings submitted until now strongly support FISCHBEIN's approach according to which the acquisition of an operative probability concept is promoted and fostered by teaching children rules and strategies of probability calculus.

Does this also mean, as BRAINERD and SIEGLER propose, that children may only cope successfully with probability tasks by means of acquired strategies, that is without knowledge? We do not think so. First of all, one has to answer the question as to how a subject working on a concrete task may accomplish the selection of an optimal rule or strategy. According to SIEG-

LER, the structure of a task plays an important part in this regard. But does this not also mean that the subject must analytically penetrate a task and generate hypotheses about its structure, for instance hypotheses about the significance of relative differences of frequencies? If this is true, in order to solve a probabilistic problem the subject must have usable conceptual knowledge. SIEGLER's finding that children skip Rule II, indeed contradicts the assumption that the subject blindly proceeds along the decision tree of his own rule-system without any conceptual knowledge. Rather, it supports the supposition that not only the selection, but also the acquisition of the rules is mediated by the acquisition of conceptual knowledge.

The results of BRAINERD's study point in the same direction. BRAINERD quite emphatically proposes that the actualization of appropriate retrieval strategies in probability judgments depends exclusively on the age-specific growth of memory space. In our opinion, however, this conclusion is too hasty. One has to take into account the fact that both the age-specific extension of memory space (CASE, 1978) and the actualization of appropriate retrieval strategies (KOBASIGAWA, 1977) are closely linked to the degree of organization and the encoding format of stored information. Both should be considered to be a function of the acquisition of semantic knowledge.

Therefore, one cannot exclude the possibility that age-specific differences in processing frequency data are also due, contrary to BRAINERD's supposition, to the corresponding differences in the recognition of specific task characteristics as being relevant for the organization of frequency data. For example, the conceptual knowledge about the difference between drawing with, and without replacement should definitely govern the selection of storing and retrieval strategies.

If our re-interpretation of BRAINERD's and SIEGLER's findings is correct, it must be concluded that both conceptual knowledge and strategic abilities are principle requirements for coping successfully with probabilistic problems. Moreover, as stated above, it has to be taken for granted that both the knowledge and the strategic components mutually influence and stimulate one another during the course of development.

Regarding the latter suggestion, there is as yet only one piece of evidence, but it is very significant. In a carefully conducted investigation NOELTING (1980) was able to show that the development of the concept of proportion may be regarded as a hierarchical process in which increasingly complex rules are established by which relations are constructed between those elements that constitute proportions. The degree of complexity of the rules which are at a child's disposal at a specific point of his/her development are supposed to depend on the particular status of his/her cognitive structures resp. conceptual knowledge, whereas, on the other hand, the application of the rules stimulates and promotes the development of conceptual knowledge.

We think that NOELTING's study marks out the basic lines of a research strategy which may help us to obtain not only more precise information about the presumed interactions between the acquisition of conceptual knowledge and strategic abilities, but also help to answer the question of at which point in the course of development the child will be able to solve which probability problems in which way. Such information would not only be of general interest for developmental psychology, but could also be applied to mathematics education.

Beyond the dispute concerning the impact of conceptual knowledge and/or strategic abilities there is a further more fundamental theoretical controversy. It concerns the question as to whether the development of probability should be conceived of as a continuous or discontinuous process. While PIAGET and INHELDER, and later HOEMANN and ROSS (1982) and others have considered this process to be a discontinuous one, FISCHBEIN proposes that the probability concept develops continuously out of the *primary intuitions* that are acquired in early childhood. These points of view seem to be irreconcilable. If, however, the concept of probability within developmental psychology is conceived to be of two kinds, just like the mathematical-philosophical concept, these opposite positions may be made compatible.

On the one side intuitive components of probability can be identified. On

the other side a concept of probability exists which is founded on explicit calculations performed by formal operations. Within the stage approach of PIAGET and INHELDER, operative cognitive structures are considered to be a necessary developmental prerequisite of the probability concept. This concept, as has been noted, is manifested in classical random situations like urns and coins. This strongly indicates that PIAGET and INHELDER have an objectivistic, calculable concept of probability in mind. FISCHBEIN's position on this issue is less clear. There can be no doubt that, like PIAGET and INHELDER, he considers the operative concept of probability taught by instructional intervention in school to be an objectivistic one. However, how the *primary probability intuitions* postulated by him are to be interpreted in this respect, remains an open question. FISCHBEIN does not explain his standpoint in this regard. He simply states: "Primary intuitions are cognitive acquisitions which are derived from the experience of the individual, without the need for any systematic instruction." (FISCHBEIN, 1975, p. 117) Similarly Hans FREUDENTHAL (1975), a famous mathematics educator, regards such intuitions to be necessary prerequisites for the constitution of such mathematical objects as the probability concept. Unfortunately, neither FISCHBEIN nor FREUDENTHAL describe the process for the emergence of *intuitive probability judgments*, nor do they explicitly define their structure.

If one considers the findings with regard to children's probability learning from which FISCHBEIN infers that there are probability intuitions in early childhood, it can be shown that these intuitions presumably are based on simple judgmental heuristics. As has been shown by WEIR (1964), younger children apply a so-called strategy of maximizing reinforcement in probability learning tasks. In other words, they prefer the response item which is most frequently associated with a contingent or non-contingent reinforcer. According to WEIR, this strategy is based on the simple hypothesis that the reinforcer is the stimulus which occurs most frequently. In our opinion this response strategy can be considered as a proto-form of a naive judgmental heuristic in the sense of KAHNEMAN and TVERSKY (1982). In a similar way the probability judgments in the spinner experiments, are based, even with older children, on area estimates (cf. HOEMANN

& ROSS, 1971 and HOEMANN & ROSS, 1982), which may be governed by simple judgmental heuristics.

Taking this conclusion to be true, the *primary probability intuitions* of younger children are to be conceived of as most likely being a kind of degree of belief and hence a *preconceptual* subjective probability concept, and not the result of a conscious reconstruction of repeated events. Consequently, the development of the probability concept may be described as a process of differentiation of progressively more elaborated *subjective probability* judgments, which are based on increasingly more complex heuristics. At the endpoint of this differentiation process the *heuristic-based* subjective probability judgments may be replaced by logical-mathematical ones, but only in situations or tasks which exhibit the above mentioned criteria of the objectivistic probability concept. It seems at least plausible that this process starts with *intuitive* probability estimates based on simple motivational, perceptual, and similar judgmental heuristics. Taking this developmental differentiation as granted, FISHBEIN's interpretation that the acquisition of the probability concept is a continuous process seems to be justified. At the same time, the opposite position taken by PIAGET and INHELDER and others is obviously due to a one-sided objectivistic interpretation of the probability concept.

Let us close this contribution with two questions that relate it more closely to the central issue of this symposium.

The first question concerns the issue: *Why and when a logical-mathematical concept of probability develops out of intuitive judgmental heuristics?*

The question as to *when* will not be answered by focussing upon a specific age period, but rather by making explicit two important developmental conditions which, in our opinion, have to precede the emergence of a logical-mathematical probability concept. The first condition is the conceptual understanding of chance and variability as well as the ability to determine and construct the event-space. The second and more fundamental condition is that of reflexivity of thinking. It is only the ability to

reflect on our own thinking and to recognize the insufficiency of our intuitive judgments that enables us to generate alternative and better strategies for handling problems. These propositions may yield an answer to the *why*. They may also have important consequences for the teaching of probability in schools. We think that shaking students naive judgmental heuristics and making them aware of their fallacies by means of suitable problem tasks within the school teaching-process may be an appropriate - and perhaps even an indispensable - intervention strategy by which the acquisition of an operative probability concept can be promoted.

The second question refers to the relationship between intuitive judgmental heuristics and the mathematical-logical probability concept, and whether there are implications for an understanding of decisions under uncertainty. The development of the logical-mathematical concept of probability - at least in western culture - is fostered predominantly, if not exclusively, by means of, and with respect to, a narrow class of situations, namely situations which are paradigmatic for the objectivist concept of probability. Restricting the application of the probability concept to such situations, does not, however, take into account all the possible recognition strategies of the subject. This means for instance, that even in situations of different types of uncertainty (cf. KAHNEMAN & TVERSKY, 1982 und HUMPHREYS & BERKELEY, in this volume), the individual is able to define elementary events, construct an event space, in which to assign probabilities to these events, and in this way build an image or model of the environment. Within this model, formal operations of the logical-mathematical concept of probability may be applied, thus rendering possible a subjective stochastical reconstruction of decisions under uncertainty.

References

- Bogartz, R.S.: 1969. Short term memory in binary prediction by children: Some stochastic information processing models. In: G.H. Bower & J.T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 3. New York: Academic Press, 299-391

- Brainerd, Ch.J.: 1981. Working memory and the developmental analysis of probability. In: *Psychological Review* 88, 463-502
- Capon, N. & Kuhn, D.: 1979. Logical reasoning in the supermarket: Adult females' use a proportional reasoning strategy in an everyday context. In: *Developmental Psychology* 15, 450-452
- Case, R.: 1978. Intellectual development from birth to adolescence: A neo-Piagetian interpretation. In: R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum, 37-71
- Fine, T.L.: 1973. *Theories of Probability. An examination of foundation.* New York: Academic Press
- Finetti de, B.: 1974. The true subjective probability problem. In: C.A. Stael von Holstein (Ed.), *The concept of probability in psychological experiments.* Dordrecht-Holland: Reidel
- Fischbein, E.: 1976. Probabilistic thinking in children and adolescents. In: IDM, *Forschung zum Prozeß des Mathematiklernens.* Reihe: Materialien und Studien Bd. 2, 23-42
- Fischbein, E.: 1975. *The intuitive sources of probabilistic thinking in children.* Dordrecht-Boston: D. Reidel Publishing Company
- Flood, M.M.: 1954. Environmental non-stationarity in a sequential decision-making experiment. In: R.M. Thrall, C.H. Coombs & R.L. Davis (Eds.), *Decision Processes.* New York: Wiley, 287-299
- Freudenthal, H.: 1975. Preface. In: E. Fischbein. *The intuitive sources of probabilistic thinking in children.* Dordrecht-Boston: D. Reidel Publishing Company
- Goodnow, J.J.: 1955. Determinants of choice-distribution in two-choice situations. *American Journal of Psychology* 68, 106-116
- Goulet, L.R. & Goodwin, K.S.: 1970. Development and choice behavior in probabilistic and problem-solving tasks. In: H.W. Reese & L.P. Lipsitt (Eds.), *Advances in child development and behavior.* New York: Academic Press, 213-254
- Green, M.G.: 1978. Structure and sequence in children's concepts of chance and probability: A replication study of Piaget and Inhelder. *Child Development* 49, 1045-1053
- Hacking, I.: 1975. *The emergence of probability.* Cambridge: Cambridge University Press
- Hoemann, H.W. & Ross, B.: 1971. Children's understanding of probability concepts. *Child Development* 42, 221-236
- Hoemann, H.W. & Ross, B.M.: 1982. Children's concept of chance and probability. In: Ch.J. Brainerd (Ed.), *Children's logical and*

mathematical cognition. New York: Springer, 93-121

Hörz, H.: 1980. Zufall - Eine philosophische Untersuchung. Bern: Akademie-Verlag

Josephson, J.: 1977. The child's use of situational and personal information in predicting the behavior of another. Unpublished doctoral dissertation, Stanford University

Kahneman, D. & Tversky, A.: 1982a. Judgment under uncertainty. In: D. Kahneman/ P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: heuristics and biases. Cambridge: Cambridge University Press, 3 - 20

Kahneman, D. & Tversky, A.: 1982b. Variants of uncertainty. In: D. Kahneman/ P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: heuristics and biases. Cambridge: Cambridge University Press, 509-520

Kobasigawa, A.: 1977. Retrieval strategies in the development of memory. In: R.V. Kail & J.W. Hagen (Eds.), Perspectives on the development of memory. Hillsdale, N.J.: Lawrence Erlbaum, 177-201

Messick, S.J. & Solley, C.M.: 1957. Probability learning in children: some exploratory studies. Journal of Genetic Psychology, Vol. 90, 23-32

Mises von, R.: 1928. Wahrscheinlichkeit, Statistik und Wahrheit. Wien: Springer

Noelting, G.: 1980a. The development of proportional reasoning and the ratio concept, part I - Differentiation of stages. Educational Studies in Mathematics 11, 217-253

Noelting, G.: 1980b. The development of proportional reasoning and the ratio concept, part II - Problem-structure and the mechanism of adaptive restructuring. Educational Studies in Mathematics 11, 331-363

Piaget, J.: 1950. Une experiment sur la psychologie du hasard chez l'enfant: La tirage au sort des couples. Acta Psychologica, Vol. 7, 323-336

Piaget, J. & Inhelder, B.: 1951. La genese de l'idee de hasard chez l'enfant. Paris: Presses Universitaires de France (1975. Translated as: The origin of the idee of chance in children. New York: Norton)

Ross, L.: 1981. The "intuitive scientist" formulation and its developmental implications. In: J.H. Flavell & L. Ross (Eds.), Social cognitive development. Frontiers and possible futures. Cambridge: Cambridge University Press, 1-42

Ross, B.M. & Hoemann, H.W.: 1975. The attainment of formal operations: A comparison of probability concepts in deaf and hearing adolescents. In: Genetic Psychology Monography 91, 61-119

Scardamalia, M.: 1977. Information processing capacity and the problem of

horizontal décalage: A demonstration using combinatorial reasoning task. In *Child Development* 48, 28-37

Siegler, R.S.: 1981. Developmental sequences within and between concept. In: *Monographs of the Society for Research in Child Development* 46, 2

Weir, M.W.: 1964. Developmental changes in problem-solving strategies. *Psychological Review* 71, 473-490

This Page Intentionally Left Blank

THE PERCEPTION OF PROBABILITY SITUATIONS IN PUPILS AGED 12 TO 14

GEORGES GLAESER

*Université Louis Pasteur
IREM de Strasbourg
France*

This is the title of a doctoral thesis submitted in Strasbourg in June 1982 by the Mexican researcher Jesus Alarcon nicknamed "Papini".

Our attention has been called to this study for two reasons:

First, the topic treated in this study concerns the teacher and the educator. This doctoral thesis investigates the (correct or false) ideas young pupils may have acquired with regard to certain phenomena of chance well before they began to study probabilities in the classroom.

When one starts to teach them this subject, they are certainly not without previous knowledge: from infancy, everybody is familiarized with situations of betting, of drawing lots or with decisions under uncertainty. The only basis for efficient teaching is the teacher's awareness of his pupils' preconceived ideas which he will seek to discourage or promote.

Developing the concept of probability is much more critical than imagined by those who already have assimilated its basic notions. Humanity has had a great difficulty in mastering the phenomena of chance and in getting to the bottom of the magical prejudices which were attached to them. The history of probability calculus has seen difficult beginnings heralded by

the names of CARDANO, PASCAL, FERMAT, d'ALEMBERT, LAPLACE, KOLMOGOROV etc... For a long time, this scientific discipline was the preferential domain of paradoxes and erroneous conclusions. And it is reasonable to think that many of the difficulties which formerly frustrated the greatest scientists must baffle our pupils still today.

Second, ALARCON has carried out his study making use of an exemplary experimental method. He refrains from formulating conclusions which are not based on established facts. He anticipates reasonable potential objections and takes experimental precautions in advance which enable him to refute these.

In my opinion, this doctoral thesis constitutes the present model of what enables us to state today that mathematical didactics has become a science.

I should like to add this in all modesty. While this thesis has been prepared under the supervision of the Strasbourg mathematics educators, ALARCON has carried it through "against" our team: the candidate always proved to be more exacting and more subtle than we were. Our role was confined to heckling him. Eventually, however, it was mostly "Papini" who influenced us!

A conflict of logics

The elementary teaching of probability generally begins with describing situations of drawing with replacement which are realized with urns containing p white balls and q black balls.

It is then postulated more or less explicitly that all answers to questions in that context depend only on the proportion $p/(p+q)$ and not on the total number $p+q$ of balls.

ALARCON, however, has established that this postulate is far from being evident to the beginner! More precisely, after this belief is established, it still remains unstable: a slight modification in the formulation of

questions will suffice to cancel it out (as it will be seen with regard to one condition of the experiment shown in fig. 2).

One result, in particular, is surprising: the persistence of answering strategies which are not based exclusively on taking into account the relationship $p/(p+q)$ is especially evident for pupils considered more advanced than the average of their peers.

In fact, beginning to learn probabilities uncovers a conflict of logics. An example: in the course of a game of heads and tails, one may bet heads. But if chance should make heads come out, nobody will consider this to be an intellectual outrage.

But if, as opposed to that, the keys of a hand-held calculator are punched $12+7$ and the result indicated is 15, everybody will be aware that something has gone wrong.

From the Cartesian point of view, it is perfectly conceivable that in the course of a bridge game, a hand of 13 spades could be dealt. From the probabilist's point of view, however, this smells of miracle or trickery.

ALARCON shows that there is an opposition between compatibilistic and probabilistic attitudes. The pupils who have adopted the former will refuse to choose between "possible" alternatives even if one is clearly more probable than the other. Those who have adopted the probabilistic attitudes will be interested in relations and consider an event whose probability is very low "impossible". (For instance those who answer "rather bag 2" or "certainly bag 2" to the question in fig 4.)

The experiment

Alarcon's study began with a pretest in which he observed about forty children of different ages in individual interviews. He submitted about thirty questions to them, adapting these, if necessary, between one interview and the next.

After this stage, he designed a questionnaire which was submitted to 6 classes of fifth-formers (12 to 13 years) and to 6 classes of forth-formers (13 to 14 years) selected at random in the French Department of Bas-Rhin (i.e. to about 300 pupils).

The questionnaire was presented in three variants ("modalités") which differed either in their formulation or in the question order.

The following is an example of a typical question that can be answered on the basis of the compatibilistic attitude.

1. Alain has taken one of the two bags shown below, and without ever looking into the bag...

Alain has drawn a ball: the ball taken out was WHITE; then he has put the ball back into the bag, mixing well.

Alain has drawn a second time: he has taken out a WHITE ball then he has put the ball back into the bag, mixing well.

Alain has drawn a third time: he has taken out a BLACK ball: then he has put the ball back into the bag, mixing well.

Alain has drawn a forth time: he has taken out a WHITE ball: then he has put the ball back into the bag, mixing well.

Alain has drawn a fifth time: he has taken out a BLACK ball: then he has put the ball back into the bag, mixing well.

Alain has drawn a sixth time: he has taken out a BLACK ball: then he has put the ball back, mixing well.

Which bag did Alain take? Place a cross in the box you prefer.



bag 1



bag 2

fig. 1



certainly
bag 1



rather
bag 1



no reason for
preferring one
of the bags



rather
bag 2



certainly
bag 2

But now consider two questions concerning the two bags below:



bag 1



bag 2

fig. 2

Suppose that a draw (with putting back) out of one of these bags be given and you are asked to indicate from which bag it is. For an average probabilist, there is no need to take note of the draw: the bags are equivalent and so there is no reason to prefer one of the bags rather than the other.

The experiment, however, yields the following result: if the draw mentioned is reduced to a single white ball, 98 percent of the (older) pupils of the fourth form and 76 percent of those of the fifth form will give the correct answer. But those percentages go down to 65 and 52, respectively, if the draw indicated is 4 white balls and 2 black balls! It seems that the pupils believe that it is easier to draw 4 white balls from

bag one than from bag two. This is the phenomenon of instability mentioned above. But this is only a local result of this thesis. In fact, the questionnaire harasses the respondent pupils with items which seem to be analogous, but really call for different strategies.

If the draw contains more white balls than black ones, the bag containing more white than black balls may be preferred without taking proportions into account. Other information is obtained on the basis of items in which the content of one of the bags is unknown. For example:

Edith has taken one of the bags shown below, then she has drawn a ball without looking into the bag; the ball taken out was WHITE.

Then she has put the ball back into the same bag.



bag 1



bag 2

fig. 3

FFV FV 10+13

FFVVV 16+14

FFVVF 7+0

Now there are five statements: which are true, and which are false? Put a cross in the respective box.

It is sure that Edith took bag 1.	true ☐	false ☐
It is impossible that Edith took bag 1.	true ☐	false ☐
It is not sure that Edith took bag 1.	true ☐	false ☐
It is more probable that Edith took bag 1.	true ☐	false ☐
It is just as probable that Edith took bag 1 as bag 2	true ☐	false ☐

Or, again:

The result of the draws was: 6 BLACK balls.

Which bag did Gerard take? Put a cross in the box you prefer.



bag 1

☐
 certainly
bag 1

☐
 rather
bag 1

☐
 no reason to
prefer any
of the bags

☐
 rather
bag 2

☐
 certainly
bag 2



bag 2

fig. 4

fourth

form	?	<u>11</u>	<u>18</u>	19	7

fifth

form	?	4	18	19	5

The method chosen by Alarcon avoids drawing conclusions from one particular answer to an isolated question. On the contrary, it shows evidence for coherent behaviour with regard to the questionnaire as a whole.

And the formulation of the questions permits to reach very accurate conclusions.

Probability and fractional calculus

Let us now see how Alarcon refutes a view very widespread among teachers and researchers in the field of education.

It would seem plausible that learning probability theory requires mastering fractional calculus, and it is common belief that this is the only indispensable prerequisite. In order to discuss this point, Alarcon presented, at the beginning of each of the three "modalités", items which were inspired by the famous orange juice test developed by Gerald Noelting.

Catherine will make up orange juice for a party.

Below each jug, there is an indication of the quantities of orange juice and water Catherine will mix in order to prepare the drink:

 = a glass of orange juice;

 = a glass of water.

Which drink will have a stronger orange taste? Put a cross in the correct box.

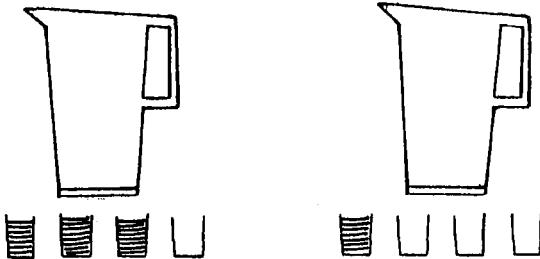
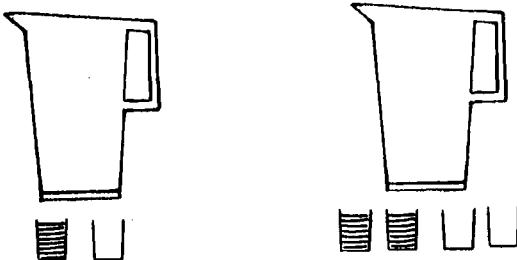


Fig. 5

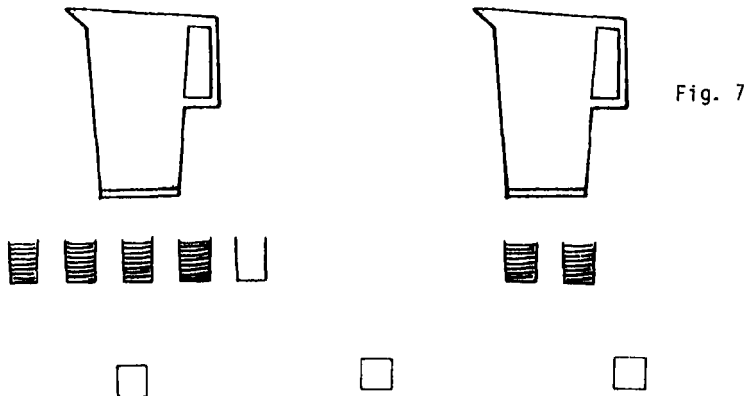


(a box for
"both drinks
taste the same"
is also
provided)

Fig. 6

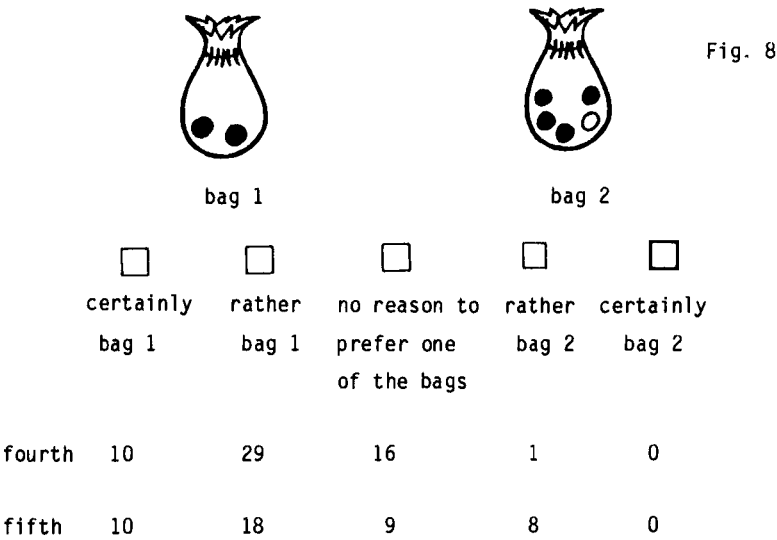
In order to permit comparisons, the experimenter has chosen the same numbers p and q of glasses of concentrated orange juice and water as black and white balls in the bags. It would seem that Noelting's experiment was a full success, while many significant failures are observed in the random ball-drawing test. The experiment furnishes a convincing experimental refutation of the claim that mastering fractional calculus is sufficient to understand the calculus of elementary probabilities.

Even better, the following test is one of those which contain a strong and unexpected proportion of failures, while the items referring to bags of the same composition are fairly successful on the whole.



same taste in the
two jugs

fourth	<u>11</u>	0	49
fifth	<u>19</u>	0	27



Development with age

As a general rule, an improvement of performance can be observed when passing from pupils of the fifth form to those of the fourth. But Alarcon is too exacting to draw conclusions if the improvements are not significant (in the statistical sense). Besides it would have been necessary, in this case, to use real age groups rather than referring exclusively to the opposition between fifth form (pupils aged 12 to 13) and fourth form (pupils aged 13 to 14).

For the cases where this progress was proved experimentally, the reader is referred to the thesis in detail.

This result, in any case, conflicts with many accepted ideas of official pedagogy: a certain maturing can be established at precise points between the ages of 12 and 14 with regard to notions which are not the object of any teaching in school.

Anyway, this summary of Alarcon's doctoral thesis is much too brief to give more than a glimpse at the rich findings of this experimental study.

All those who should like to get a more profound view of the opportunities offered by the experimental method in educational research are advised to read Jesus ALARCON's thesis carefully.

This Page Intentionally Left Blank

THE PSYCHOLOGY OF INFERENCE AND THE TEACHING OF PROBABILITY AND STATISTICS: TWO SIDES OF THE SAME COIN ?

J. MICHAEL SHAUGHNESSY

San Diego State University
Department of Mathematical Sciences
USA

For many years teachers of probability and statistics have been struggling to improve the plight of the naive statistician, the student in an introductory course in probability and statistics. Such students may have a first encounter with statistics either at the secondary level or in college, and are probably poorly prepared mathematically for their plunge into this new subject. They may believe that all outcomes from an experiment are equally likely, for example that the chance of getting a sum of 4 on a roll of two dice is the same as that of getting a sum of 7. They have little experience with the rapid growth and decay of partial products, and so they make very low estimates for the number of different possible batting orders for a nine-person baseball team, and give very high estimates for the number of people needed to have an even chance of matching birthdays. Indeed, the naive student of probability and statistics possesses a myriad of mathematical misconceptions which present a formidable challenge to the teacher of statistics.

Within the last ten years a considerable amount of investigation by cognitive and social psychologists in the area of judgment and decision making has produced evidence which suggests that many misconceptions of probability and statistics have psychological roots. Mere exposure to the theoretical laws of probability and statistics may not be sufficient to overcome probabilistic misconceptions (TVERSKY and KAHNEMAN, 1971). Thus we are led to consider the naive teacher of statistics. Students show up in an introductory statistics course equipped with a formidable array of mathematical misconcep-

tions and deficiencies. On top of that, we find out, these same students use intuitive heuristics such as availability and representativeness when estimating the likelihood of events. They ignore base rates when making predictions. They are slow to change opinions in the face of new data and they are poor at detecting covariation. They may not even "believe" in probability. In short, they are a statistical disaster. If the students' misconceptions aren't enough to deter our naive teachers of statistics, they can be comforted by the fact that *they too* are subject to the same psychological biases and misconceptions as their students.

The student is caught in the middle as the judge and decision maker of the future, the future user (or non-user) of statistics. How can research in the learning of statistics and probability help the student as naive judge? How can research in judgment and decision making help the student as naive statistician? How can research from both sides of the coin help the teacher of statistics? These questions are posed from the perspective of mathematics education. Implicit in the above questions are the beliefs that 1) it is possible to improve people's understanding of probability and statistics; 2) it is possible to improve people's judgment and decision-making processes; 3) items 1) and 2) are intricately connected with each other, and so they should be connected in the curriculum.

In this paper we will explore the interrelationship of judgment and decision making with probability and statistics in three parts. Part one will examine research on obstacles to the use of statistical principles when making inferences from the perspectives of psychology and mathematics education. Part two will review some suggestions from psychologists and mathematics educators for increasing people's knowledge and reliance upon statistics when making inferences. Finally, part three consists of some recommendations for further research.

1. Decision Making Under Uncertainty and the Use of Statistics

Reflections from Psychology

It is certainly not the intent of this paper to engage in the monumental task of surveying ten years of literature on judgment and decision making. Several excellent reviews and syntheses of this literature have recently been published (SLOVIC, FISCHHOFF, and LICHTENSTEIN 1977; NISBETT and ROSS, 1980;

KAHNEMAN, SLOVIC, and TVERSKY, 1982). However there are aspects of this research that may be of particular interest to researchers in mathematics education. We shall touch upon several of these aspects briefly.

The use of heuristics. Considerable evidence has been gathered to show that people tend to rely upon simplifying heuristics when making estimates for the likelihood of events. (KAHNEMAN and TVERSKY, 1972, 1973a, 1973b, 1974). People's use of such heuristics as representativeness, anchoring, and availability to make inferences has been replicated and extended many times since these initial studies were published. (BAR-HILLEL, 1982; TAYLOR, 1982, SHAUGHNESSY, 1977). The use of these heuristics has been found not only among statistically-naïve subjects but also among supposed statistical experts (TVERSKY and KAHNEMAN, 1971). Although these seat-of-the-pants heuristics often guide one adequately in daily affairs, they can be misleading, particularly when inferences are to be made from statistical data. Thus, people are insensitive to the effects of sample size on predictive validity, unaware that data tend to regress to the mean and may be overconfident of prediction based upon invalid input data if they employ representativeness. They may tend to rely on vivid anecdotal data rather than carefully collected statistics if they employ availability. They may "lock in" to an initial estimate and be slow to revise when faced with new data if they employ an anchor. The existence and prevalence of these heuristics poses added challenges to the teacher of probability and statistics.

The disregard for base-rate data. In many situations when people must combine several pieces of information to make an estimate for the likelihood of an event, they tend to ignore base-rate data. This behavior is particularly evident in many situations where an application of Bayes law would give a good normative prediction. Research on problems such as the taxicab problem (TVERSKY and KAHNEMAN, 1980) has created a flurry of theoretical activity in an attempt to explain why people attend to individuating data and ignore base-rate data. Here is one version of the taxicab problem which has been used by many researchers.

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- (i) 85% of the cabs in the city are Green and 15% are Blue.
- (ii) A witness identified the cab as a Blue cab. The court tested his

ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability that the cab involved in the accident was Blue rather than Green?

Results show that people attend to the witness reliability information and ignore the base-rate data. Estimates for the likelihood of the blue cab are, thus, too high. AJZEN (1977) suggests that when base rates have causal implications for outcomes, people's predictions *will* be influenced by them. NISBETT and BOGIDA (1975) and NISBETT, BORGIDA, CRANDALL, and REED (1976) substantiate the base-rate phenomenon when people are given consensus information and target specific information. They suggest that people ignore the consensus (base-rate) information because of its abstract nature and attend rather to the concrete actor or target related information which is much more vivid. They raise concern that people make errors in predicting a whole event from its biased parts, or conversely from inferring parts or instances from a whole. They are also concerned that ignorance of base-rate data could have serious implications for the teaching of social psychology, and refer to the "Oh, that could never happen to me" phenomenon in their students. (We suggest that it also has serious implications for the teaching of probability and statistics). BAR-HILLEL (1980) found that base-rate information will be combined with other information if the two kinds of information are perceived as equally relevant to the task. Although the very nature of the task itself may have some bearing on how base rates are used (AJZEN 1977, BAR-HILLEL 1980). LYON and SLOVIC, (1976) found the base-rate phenomenon to be highly robust across a wide variety of tasks. FISCHHOFF, SLOVIC and LICHTENSTEIN (1979) found that changes in base-rate information did result in subjects shifting their predictions in the direction that normative theory would suggest, however, insufficiently. Corresponding changes in sample size did not produce such shifts.

For the mathematics educator and teacher of statistics, there is much to be concerned about with the ignorance of base rates and insensitivity to sample size. Somehow base-rate information gets lost in the shuffle, and people may or may not utilize it depending on vividness, saliency, causality, or perhaps knowledge about probability and statistics itself, and the relevance of that knowledge to the situation.

Unwarranted confidence in precision. When confronted with new data or information that conflicts with beliefs, people are slow to change inferences or decisions. NISBETT and ROSS (1980) reviewed and summarized the literature on theory maintenance and change (see Chapter 8), and concluded that people tend to adhere to theories that are far from normatively justifiable. Initial evidence seems to carry more weight than subsequent evidence. When presented with conflicting evidence, subjects tend to polarize inferences and base them on *a priori* belief systems (witness the effects of reports of the Vietnam War on the American people). People commit prematurely to a position, and revise slowly from a position. Apparently people tenaciously maintain their own microparadigms (KUHN 1962) which are as resistant to change in individuals as the scientific paradigms are to change in scientific theory. So, the teacher of probability and statistics not only has psychological biases and mathematical misconceptions to contend with, but also personal belief systems among students.

There are a variety of individual inference strategies. Recent research by WELL, POLLATSEK, and KONOLD (1982) suggests that the types of misconceptions that occur among naive subjects when confronted with Bayesian problems, such as the taxicab problem, are much more complicated and varied than simple neglect of base-rate data or use of some causal schemas. Much of the research on misconceptions of probability has reported only average responses--means or medians--and has not asked the students *why* they responded in a particular way. WELL carried out follow-up clinical interviews with students who had been given Bayesian problems, like the taxicab problem, in order to disclose their mental processes. He found large individual differences in the types of strategies employed and in the kind of misconceptions that were apparent among the students. In particular, he found little evidence to support the causal schemas contention of TVERSKY and KAHNEMAN (1980). Rather, some students used non-probabilistic strategies to estimate the likelihood of events, strategies which may be based on a deterministic model or even on personal belief systems. WELL says, "It is difficult to infer a strategy from a numerical answer. We have observed subjects with markedly different strategies (in the interviews) arrive at the same numerical answer, and subjects with the same strategy arrive at different answers."

An estimate for the likelihood of an event may not be based upon a probabilistic strategy at all. For example, suppose someone decides that the

witness *is* right in the cab problem. They will then conclude that the blue cab was involved in the accident even though the base rate is green-cab dominated. Chance is not involved in such decisions. In the interviews WELL and his colleagues obtained estimates as high as 90-95% that it was a blue cab. When asked why, these subjects said, "I think the witness was right," or "I think it *was* a blue cab." In such cases, the decision may be based on a belief or deterministic decision. The percentage estimate given by the subject is more like a confidence rating than a probability estimate. Base-rate information may be neglected because it violates preconceptions or beliefs of the subject.

This type of misconception may pose the most difficult challenge of all to those who teach probability and statistics. What do you do if your students just plain *don't believe in probability* or chance? WELL et. al. (1982) reports a protocol excerpt in which a college student is asked to pick which side he thinks will turn up most often if an irregularly-shaped die (actually a bone) is rolled. During the interchange the student eventually settles on side C as the most likely to come up--for no apparent probabilistic reason. (In fact, there is no evidence of an understanding of chance by the student during the entire episode.) The student is then shown data from 1000 rolls of the die which clearly suggest side D is most likely to occur, and then asked which would be the most likely. "You found D most likely. ...I still think C." This student has evidently formed a belief that C *is* most likely, and experimental data are not going to alter his belief.

WELL's research suggests caution in attributing inference strategies such as representativeness or availability, or neglect of base-rate data (for whatever reason) to individuals based upon summary numerical results. There may indeed be *many different strategies* employed by the individuals which are lost in summary type statistics. Clinical interviews of subjects' reasoning processes provide an alternative methodology to investigate strategies. Several authors have suggested that accuracy of subjects' introspective reports is in serious question when they make judgments. (NISBETT and BELLOW, 1977; NISBETT and WILSON, 1977). Subjects appear to be unaware that their judgments were affected by certain stimuli, and thus are perhaps unreliable in their introspective report. However, much use has been made of "thinking aloud" techniques in the field of problem solving in recent years (NEWELL and SIMON, 1972; SCHOENFELD, 1979). WELL and his colleagues posed the taxicab problem

to their subjects as a problem to be solved, rather than as a judgment to be made. In the context of problem solving, introspection may indeed reliably detect the strategies employed by subjects.

Reflections from Mathematical Education

The psychological literature of judgment and decision making has primarily used adult subjects, often the "freshman psych major" population, for its investigation. In contrast, much research in mathematics education has dealt with young children's intuitions of probability. Perhaps this is in part due to the relatively recent translation into English of PIAGET's (1975) work on the origin of chance in children.

FISCHBEIN (1982, 1976, 1975; FISCHBEIN, PAMPU, and MANZAT, 1970a, 1970b) has investigated young children's conception of chance and also the effects of teaching on their combinational and probabilistics ability. He concluded (1976) that children and adolescents do not distinguish the necessary from the possible in many situations, and thus are ill-equipped to reason probabilistically. According to FISCHBEIN, this could be an artifact of our instruction in science education which has emphasized deterministic models at the expense of the study of uncertainty. Indeed, almost all of our mathematics instruction is based upon deterministic models. Consider Euclidean Geometry, for instance, or Newtonian Physics. Once the laws or postulates are stated, everything else is determined, and there is no room for "chance." It may not be so surprising that people have so many misconceptions of chance and probability, if one looks at their educational background.

Several studies have reported on misconceptions of probability in older subjects. In order to survey the intuitions of chance and probability concepts of adolescents in England, GREEN (1982) tested four thousand 11- to 16-year-old secondary students from the midland counties. After frequent piloting, he developed a probability concepts text to examine students' notion of chance, randomness, probability as ratio, expected value, and also their verbal understanding and use of such phrases as "equally likely," "unlikely," and "influenced by chance." Here are some of the conclusions listed by GREEN which may be of interest to both psychologists and mathematics educators:

- a) Items requiring ratio concept are very poorly done.
- b) Various strategies are employed by pupils when answering odds problems, and there is little consistency in choice of strategy either

by individuals or by classes.

- c) Items involving an appreciation of randomness, the stability of frequencies, and inferences were particularly poorly done, with *little improvement with age*.
- d) Pupils' verbal ability is often inadequate for accurately describing probabilistic situations.
- e) Very often, certainty and high probability are equated, as are impossibility and low probability.
- f) Most pupils do not attain the level of formal operations in probabilistic thinking before they leave secondary school.

GREEN's results with secondary students substantiate some of the findings of the psychological literature. In particular, the students had little notion of randomness or inference. Also, the issue of determinism is once again raised as GREEN's subjects felt the "very likely" was necessary.

Evidence for misconceptions of randomness among older subjects has been documented by FALK (1981a). Subjects were shown sequences of 21 cards of two colors, say 10 green and 11 yellow, and were asked to rate the sequences for randomness. Subjects committed both type I and type II errors, seeing patterns where none existed and inferring randomness when a non-chance sequence was to be judged. FALK concludes that quite often people reject the chance hypothesis when making judgments and search for alternative explanations. In another report FALK (1981b) points out that "miracles" and "coincidences," such as chance meetings between friends, are much more likely to occur than we think. People tend to look at the *intersection* of all the minute details that surround a coincidence, and thus judge it to have a very low probability of occurring. If events surrounding a so-called coincidence are viewed *disjunctively* instead of *conjunctively*, it is possible to demonstrate mathematically that the event is much more likely to occur than we imagine.

The studies of FISCHBEIN, GREEN, and FALK suggest that, at least in part, people's misconceptions of probability may be due to a lack of experience in probabilistic thinking, a lack of exposure to probabilistic models in our schools, and an overemphasis upon deterministic reasoning in our schools and society. KAHNEMAN and TVERSKY (1971) have voiced concern that even people trained in probability and statistics still are subject to the same biases and misconceptions as naive subjects, and so perhaps mere exposure to the laws and principles of probability and statistics will not change people's manner of

making inferences and judgments. However, let us consider *how* a student usually encounters probability and statistics in, say, a college course. Probability may be introduced through sets, sample space, events, and formal counting principles. The model of teaching is most likely a formal lecture where the student watches Professor X "solve problems." In addition to the formal lectures and the formal mathematized model of probability, the student has to cope with a brand new formal language--combination, permutation, event, sample space, and so on. First encounters with statistics are probably no less formal. The student is presented with some "data" and told that there are some important calculations to be done with these numbers--mean, variation, standard deviation--and that he had better know them. Why should students be good at making inferences if the importance of using data to make inferences is never mentioned in their course work or in their experience? To borrow a phrase from Nisbett, what may be needed here is a "vivid, concrete" presentation of data to focus the attention of students on the importance and usefulness of statistics.

In an attempt to overcome naive college students' reliance upon the heuristics of availability and representativeness, SHAUGHNESSY (1977) developed an experimental course in probability and statistics. The course consisted of a series of activities which emphasized experiments, data collection, simulation of experiments, making inferences from experimental data, model building, and comparing theoretical results with experimental results. The course was conducted with statistically-naive college freshmen (mostly) and met for ten weeks, five times a week for an hour. The activities have been described elsewhere (SHAUGHNESSY 1977, 1982). Students worked in small groups of three or four, made preliminary guesses for the likelihood of events, carried out an experiment and then tabulated, organized, and analyzed their data to make the inferences requested in the activities. There was no traditional lecturing in the experimental sections of the course, as the instructors were primarily facilitators for the activity in the small groups. Several sections of the experimental course were compared to sections of a more traditional lecture course on probability and statistics, to test for student reliance upon the heuristics of availability and representativeness, both before and after exposure to probability. The results of the subjects' performance on some of the early tasks of KAHNEMAN and TVERSKY (1972, 1973) indicate that students in the small group activity-based course relied less on representativeness and availability than students in the usual lecture approach.

II. *Some Suggestions for Improving Reliance upon Statistics when Making Inferences*

Suggestions From Psychology

There have been several viewpoints expressed by psychologists on the possibility of improving judgment and decision making by education in statistics. We have already mentioned the skeptical results of KAHNEMAN and TVERSKY (1971) who found biases and misconceptions among their statistically well-trained colleagues. SLOVIC, FISCHHOFF, and LICHTENSTEIN (1976, 1977, 1980) suggest that humans may be primarily trial-and-error learners when it comes to making decisions and thus suffer the consequences. We may lack adequate probabilistic cognitive processes to deal with decision making, a great danger in this nuclear age. They suggest that both the experts and the public need to be educated in the subtleties of statistical thinking starting as early as the lower grades in school. They do not, however, hold out much hope for the effects of education on people's judgmental ability. People tend to be overconfident of their ability and possess beliefs that are extraordinarily resistant to change. Furthermore, education may just make people aware of "more things that can go wrong."

NISBETT and ROSS (1980) hold out hope that people's inference errors are more frequently cognitive than emotional (also reported in LA BECQUE, 1980). They claim that the emotional and psycho-analytic aspects of decision making have perhaps received over-emphasis. They, too, recommend the early teaching of probability and statistics. Secondary and college courses in probability should provide a better grounding in inferential rules that emphasize the contrast between formal and informal intuitive strategies. Concrete examples of intuitive errors should be given in such courses. At the end of their book (1980) NISBETT and ROSS outline what they refer to as the "fortune cookie didactic" which should be employed when making decisions. Roughly, the fortune cookie didactic suggests that we:

- 1) Look at judgmental questions *empirically*--obtain and analyze data.
- 2) Pay close attention to how samples are obtained--which hat did they come from?
- 3) Check for covariation--what do the other three cells look like?
- 4) Beware of the fundamental attribution error when making judgments or inferences--behavior may be due to situations rather than personal disposition.

5) Since exceptions to our personal theories are too easily explained away, use new data, especially if it persists and is replicable. The fortune cookie didactic contains recommendations that are reminiscent of those found in Darrel Huff's book *How to Lie with Statistics* (1954).

Although NISBETT and ROSS suggest cognitive approaches to improving decision making, NISBETT, BORGIDA, CRANDALL, and REED (1976) remind us that people are not moved by dry statistical data. Effective, concrete, emotionally interesting ways of presenting data need to be used to persuade people to pay attention. They provide several anecdotes at the end of their article to demonstrate the effect of vivid concrete information on decision making. Included are references to the huge increase in visits to cancer detecting agencies since former first lady Betty Ford's operation, and the effect of live television coverage of the Vietnam war on the American public. Perhaps both hot and cold cognitions--emotional and cognitive processes--are essential to the improvement of judgment and decision making. In order to get a judge to *use* statistical data, one first has to get him to *attend* to it. (One cannot help but recall the anecdote of the farmer who helped his neighbor to get his donkey moving. The farmer watched for a while as no amount of coaxing and yelling by his neighbor could get the donkey going. Finally, the farmer picked up a board and hit the donkey over the head. "You have to get his attention first," he said. It may be quite the same with people and their use of probability when making inferences.)

Suggestions From Mathematics Education

During the past several years, there has been a rekindling of concern about the teaching of probability and statistics at all school levels, elementary, secondary, and introductory college. This concern for statistical literacy has up until now not been much influenced by research results from psychology. Rather, there has been a growing concern in the mathematical education community for improving the statistical education of our students who must face an increasingly technological job market, and become future consumers who are confronted with masses of "statistical" data in newspaper, radio, and television. Unfortunately, the teaching of probability and statistics in the schools has taken a back seat to other topics in the curriculum for quite some time. This is quite apparent in the United States, where students tend to either drop out of mathematics after a year of algebra and a year of geometry, or to follow the "college prep" track and take a second year of algebra and a year of calculus and analytic geometry. This curriculum,

which has been entrenched for many years in schools in the states, leaves very little space and time for topics from probability and statistics. Some recent developments have attempted to refocus the attention of mathematics educators upon the importance of probability and statistics. The 1981 Yearbook of the National Council of Teachers of Mathematics was devoted to the teaching of Probability and Statistics (SHULTE, 1981). This book contains a number of interesting approaches to the teaching of statistics which are appropriate for grades 5-12. Emphasis is placed upon experiments and simulations.

A series of exploratory activities called *Statistics and Information Organization* (Mathematics Resource Project, 1977) was developed in the United States for middle school students (grades 5-8). At the same time a project in Great Britain has been developing topical modules for teaching statistics to secondary students (HOLMES and TURNER, 1981). Indeed, it is not for want of excellent materials that probability and statistics is slighted in the schools. The problem is to disseminate these materials and to convince more schools that teaching probability and statistics is important. As a first step toward this goal, the First International Conference on the Teaching of Statistics (ICOTS I) was held in Sheffield in the summer of 1982. Sessions at this conference included papers on intuitive notions of chance among children and adolescents.

Several authors have made specific recommendations for the teaching of probability and statistics in order to improve statistical literacy and confront misconceptions of probability that are of a psychological nature. After investigating the probabilistic concepts of 11- to 16-year-olds GREEN (1982) gave a number of recommendations for the teaching of probability and statistics in the schools. Among his suggestions were:

- 1) There is a great need for *practical activity* at an early age, emphasizing guided discovery.
- 2) Emphasis should be placed upon mathematical modelling and interpretation in school mathematics.
- 3) Ratio is a prerequisite of many complex probabilistic situations, and thus more emphasis should be placed upon students' understanding of fractions and ratio.
- 4) The concept of randomness is sorely neglected in our deterministically oriented educational system. This is unreasonable given that we are surrounded by so many random occurrences.
- 5) Students have no experience with drawing inferences from data, and

this requires immediate attention in the curriculum.

STEINBRING and VON HARTEN (1982) suggest that statistics be introduced via the concept of relative frequency, starting with applications. They also promote multiple approaches to the teaching of stochastics, both objective (frequentist) and subjective (Bayesian). According to them, stochastics at the school level should emphasize both the transformation of data into insights and the theory of decisions. FALK (1980) suggests that we continually carry out subjective sensitivity analysis with our students in the classroom, changing the value of parameters in order to help students improve their intuitions as to how data can affect inferences.

SHAUGHNESSY (1981, 1982) has set forth a number of suggestions for teaching introductory probability and statistics to secondary or college students. These include:

1. *Teach from guesses, to experiments, to simulation, to theory.*

Introductory probability and statistics should be activity-based and experimental. We should proceed from guesses to experiments to simulations of experiments, and only after considerable intuition has been developed, to a formal mathematical presentation. We can confront students with problems in which misconceptions arise. Results from a previous study (SHAUGHNESSY, 1977) suggest that students have a better chance of overcoming misconceptions of probability that are due to intuitive heuristics such as representativeness or availability if they are introduced to probability in experimental activity-based mode rather than in a formal lecture mode. Students can work in small groups, make guesses for the likelihood of events, perform an experiment, organize and analyze the data, and make estimates for the likelihood of events based on their data. More advanced students can perhaps attempt to build a theoretical probability model for the experiment and then compare theoretical predictions to experimental outcomes. The reasons for any discrepancies between experimental results and theoretical predictions, such as sources of bias in the data-gathering process, can be discussed. An experimental approach to probability may, in fact, be the only way to break the "belief barrier" for those students who do not understand chance at all, and who rely upon causal schemas or other deterministic explanations of probabilistic phenomena.

2. *Emphasize the confounding role that belief systems can play in judgment and decision making.* At the same time, teach the importance and usefulness of statistics for helping to make objective judgments and decisions.

Suppose that you are taking a statistics course which has weekly quizzes with 10 questions each. Your average all year has been 7 correct out of 10. For the next quiz, you study with a friend and get 9 out of 10. Did studying with a friend help?

One possible response to this question is: "Of course it helped, isn't 9 better than 7?" Another possible response is: "Probably not, 9 could have just happened that time since your average of 7 is close to 9." If you ask this question of students in an introductory course in statistics, you will get both the answers above, in addition to a spectrum of "it depends" answers in between. These responses rarely indicate the need for any sort of calculation. In order to resolve the argument in an *objective* way, the class needs some data. This is an excellent opportunity to carry out a simulation, say with random numbers, and see just how likely it is to obtain the event 9 out of 10 correct if you have probability of success $P(s) = .7$ on any one item. (It is not necessary--or advisable in this author's opinion--to do a calculation using the binomial distribution the first time a student sees this type of problem.)

Suppose that a consumers' group feels that the gas tanks on a certain model of automobile are likely to explode on impact from the rear. This group has been formed because of the unfortunate death of a friend in just such an accident. The consumers' group has formed a *belief* about a car model. They may even have found other instances of such an accident. Should the car be recalled?

We must attempt to make an objective decision, if possible. Clearly there is some need for more data, and a class in introductory statistics could perhaps be led to suggest some reasonable experiment that can be carried out to determine the safety of the gas tanks. In a problem such as this, the importance of sampling in making decisions from statistics can also be brought out. (Would we do our experiments on cars all from one plant? Would we smash all the cars?)

If we focus on the usefulness and importance of statistics for making objective decisions, we may be able to temper the effects of belief systems or deterministic approaches to probability in our students.

3. *Sensitize students to misuses of statistics.*

If we demonstrate the importance of statistics in making objective decisions we are also obliged to show our students how statistics can be used to mislead or to "influence" our belief systems. Books such as *How to Lie with Statistics* (HUFF, 1954) should be required reading in all introductory pro-

bability and statistics courses. Such books can sensitize our students to misuses of statistics, deliberate or otherwise, and encourage them to analyze probabilistic statements critically.

Students can be asked to keep a log of misuses of statistics that they discover. Articles can be taken from newspapers or periodicals and analyzed for correct or incorrect use of statistics. Students who have kept such a log have ferreted out misleading graphs, inflated percentages, biased samples, insufficient sample size, and verbose statistical quantitative descriptions that had little or no basis in fact. One student discovered an article on the front page of the largest-selling newspaper in a major U.S. city that mentioned that 95 percent of the people who were interviewed were against cross-district busing to achieve racial integration in the schools. A continuation of the article (on page 14) mentioned that 70 percent of the people interviewed did not have children in the schools.

Even "literate" newspaper columnists are prone to hurtle statistics about in defense of their position, with total disregard for the data base. In an article celebrating a new high school algebra textbook for its successful battle against algebraic illiteracy, William F. BUCKLEY (1982) tosses percentages about with no reference to the original data, and perhaps little basis in fact. He claims first-year algebra students using this new text outscored second-year algebra students "by an astonishing 200%." We should encourage our students to be very critical of such loose floating percentages. Teaching the misuses of statistics is a very *concrete, vivid* (NISBETT et. al., 1976) approach which emphasizes the value of statistics.

4. *Give students challenge problems which may require them to gather or simulate their own data, and then make a decision.*

After several weeks in an introductory probability and statistics course, challenge problems can be given which confront misconceptions of statistics, demonstrate practical usefulness, and encourage simulation of experiments. Here is a sample of such problems given by the author in the past. Sources for the problems include MOSTELLER (1962), MOSTELLER et. al. (1973), and RADE (1977), as well as some problems from the literature of judgment and decision making, and some created by the author.

Stimulation by Simulation

The questions in *Stimulation by Simulation* are somewhat open-ended in nature. In many cases it is necessary to first define clearly what the

problem is, and then to adopt a mode of simulation which will enable you to "model" an experiment which may be too cumbersome to actually carry out.

Before doing any of these problems, *always write down your best guess first*, and only then carry out a simulation of the problem.

A simulation of probability experiment involves:

- a) modeling an experiment by using some apparatus with known probabilities (i.e., coins, dice, spinners, random numbers)
- b) "performing" the experiment *many* times with your apparatus, thus pooling results of several small groups may help to get a "large enough" sample.
- c) gathering, organizing, and analyzing the data (some statistical skills)
- d) calculating experimental probabilities, or other experimental outcomes (i.e., frequencies) from the data
- e) making inferences or drawing conclusions from the experimental outcomes, that is, looking back.

Problems for Stimulation by Simulation

1. A cereal company has put plastic models of 5 Star Wars characters in boxes of SPACY-O's. The company puts one character in each box. How many would you expect to have to buy in order to get all five characters? (Write down a guess first.)
2. Five friends are sitting around a table at a party discussing the merits of astrology. What is the probability that at least two of them have the same astrological sign? (Assume the signs are equal in duration.)
3. Some data were obtained from the records of the Bureau of Meteorology in Perth, Australia. A sunny day was defined as a day on which the officially recorded number of hours of sunshine was 70% or more of the maximum possible hours of sunshine for that day of the year. The months April to September in 1973 seemed to be typical months in terms of average number of hours of sunshine. A classification of each day in this period as "sunny" or "dull" (S or D) and recording of the events led to the following:

D followed by D: 79 occasions
 D followed by S: 33 occasions
 S followed by D: 33 occasions
 S followed by S: 37 occasions

Thus the proportion of dull days followed by dull days was $79/112 = .70$, and the proportion of sunny days followed by sunny days was $37/70 = .53$.

- a) What is the "average" length of runs of consecutive dull days?
 - b) What is the "average" length of runs of consecutive sunny days?
4. Two points (numbers) between 0 and 1 are chosen at random. The

points divide the interval from 0 to 1 into three line segments. What is the probability that the three segments will form a triangle?

5. About one-half the babies born are boys. Would there be more days during the calendar year when at least 60% of the babies born were boys:
 - i) in a small hospital
 - ii) in a large hospital
 - iii) makes no difference
6. Suppose you are playing a game in which you have a 50% chance of winning. Bet a dollar. Each time you lose, double your bet. Each time you win, go back to betting a dollar.

Is this a good betting system?
7. Fast Al suggests to Slow Sam that they play the following game. They will each throw either one or two fingers. If Sam throws one and Al throws two, Sam wins \$30. If Sam throws two and Al throws one, Sam wins \$10. If they throw the same number of fingers, Al wins \$20.
 - a) If each player plays a 50-50 strategy between their two choices, who wins in the long run?
 - b) What is the best strategy for Slow Sam, i.e., how should he divide his choices?
8. From where he stands, one step toward a cliff would send the drunken man over the cliff. He takes random steps, either towards the cliff or away from it. At any step, the probability that he takes a step toward the cliff is $\frac{1}{3}$, and the probability that he steps away from it is $\frac{2}{3}$. What is his chance of escaping a fall?
9. 16 teams are entered in a double elimination softball tournament. What is the expected number of games a team will play:
 - i) if the team has a 50% chance of winning each game?
 - ii) if the team has a 70% chance of winning each game?
10. Several years ago a certain beer company claimed that they "tried their famous High Life taste test on many a beer expert, and no one had yet correctly identified which beer came from a can, which from a bottle, and which from a tap."

How likely is it that the beer company was telling the truth? Or, perhaps another way to put it, how many beer experts do you suppose the company really tested?
11. Suppose a man faces north and tosses a coin to decide whether to take one step north or one step south. The man does this for twenty-five tries.

- i) On the average, how far is he from the starting point?
 - ii) On how many steps is he on the north side of his starting point?
On the south side?
 - iii) How often does he return to the starting point?
12. Three towers are built of 8, 6, and 4 blocks. A tower is chosen at random and a block removed.
- a) What is the expected number of blocks that will be removed before one of the towers has been "leveled" (i.e., no blocks left)?
 - b) What is the probability the 4-block tower will be destroyed?
 - c) Can you think of a "real world" situation for which this tower game is a model?
13. (The big-bad-triangle, or three's a crowd.) Three duelers, the Good, the Bad, and the Ugly, are to fight a three-cornered pistol duel. All know that the Ugly's chance of hitting his target is .30, the Bad's is .50, and the Good (of course!) never misses. They are to fire at their choice of target in succession in the order Ugly, Good, Bad, cyclically until only one man is left unhit. A hit man loses further turns, and is no longer shot at. Ugly has the first shot. What should he do to have the best chance of survival? (Running away doesn't count!)

5. *Expose Students to Statistical Paradoxes*

We have already mentioned the importance of sensitizing students to the misuses of statistics and emphasizing the role of statistics in making judgments and decisions. There are times when the way we choose to analyze data, or the underlying assumptions we make in our statistical model, will influence judgments or decisions. It is very important that we expose our students to instances of these so-called "paradoxes" so that they remain highly critical of interpretation of data when reading results. If there is more than one possible interpretation for the data, we should encourage our students to examine both interpretations. Several recent articles present instances of so-called paradoxes that are accessible to secondary school students in an introductory probability and statistics class (REINHARDT 1981; WAGNER 1981). How, for example, is it possible for baseball player A to have a better batting average than player B *both* against right-handed pitching and against left-handed pitching, and yet player B gets to start games because she has a better *overall* average than player A? This example, taken from REINHARDT's article, points out the importance of a weighted average, and the resulting

possible dilemma. FALK and BAR-HILLEL (1980) mentioned how weighted average leads to the paradox that the average IQ in all 50 U.S. states can be raised merely by redistributing the population.

Another paradox presented in both the REINHARDT and WAGNER articles involves the conviction of a couple in a court case based upon six characteristics the witnesses testified were present in the couple at the scene of the crime. The prosecution computed the odds as 1 in 12 million that these characteristics would simultaneously occur, and obtained a conviction based primarily on this calculation. The verdict was subsequently overturned when a mathematician explained that the six characteristics mentioned by the witnesses were far from being clear independent events, and that in fact one could probably find quite a few couples near the scene of the crime with the same six characteristics (Time, 1968). The *role of assumptions* in our interpretive models is crucial, and we should impress this upon our students. Indeed, coincidences may not be as coincidental as they appear to us (FALK, 1981b).

6. *We should expose both teachers and students of probability and statistics to the research results in the literature of judgment and decision making.*

It is the author's experience that very few statisticians or teachers of probability and statistics are even aware of the literature of judgment and decision making, or of the implications of this research for teaching. It is not enough to look at the student from only one side of the coin, and decide, "They are just statistically ignorant. We have only to teach them these things." Our students also have intuitive heuristics and schemas for dealing with inferences, and many of these are psychologically based. NISBETT and ROSS (1980) report that a set of statistical notes is being developed at the University of Michigan which emphasizes judgmental heuristics and the conflict between intuitive and formal statistics. This approach should become much more widespread (DOUMOUCHEL and KRANTZ, 1979).

III. *Some Suggestions for Further Research*

Researchers in judgment and decision making are concerned about understanding and improving the process of human inference under uncertainty. On the other hand, researchers and teachers in statistical education are concerned about the level of statistical illiteracy that is so prevalent among our school children, college students, and the general public. This

suggests that cooperative research efforts between the two disciplines may help both sides of the issue. Here are some questions that warrant further investigation.

1. How much of the problem of reliance upon heuristics and biases is due to ignorance of the subject of statistics and inexperience with randomness, and how much is due to psychologically-based schemas or scripts (ABELSON 1976) that would be beyond change by reasonable educational efforts? SCHOLZ (1981), for example, has conducted an analysis of nineteen different stochastic problems from both a psychological and a teaching point of view. Such studies are essential in order to separate psychological intuitions from lack of background knowledge as contributors to poor inferences among humans. Those who research from the point of view of statistical education should pay attention to the psychological results, and incorporate judgment and decision-making tasks into their research designs. On the other hand, psychologists could gather base-rate data on the level of statistical competence of their subjects, and control for that variable in their investigations. Specific questions which should be investigated cooperatively include:

a) Can instruction which is aimed at integration of base-rate information overcome people's tendency to ignore base-rate data? For example, would the introduction of statistics via random variables simulated by spinners--as developed by BENZ (1982)--improve people's ability to combine information in judgmental tasks? Would the introduction of Bayes Theorem first through simulation experiments and then through 2x2 contingency tables with relative frequencies help people to focus on the *restricted* sample space which plays so vital a role in conditional probability problems? That is, suppose we had our students translate percentages or probabilities into relative frequencies in problems such as the cab problem. We could focus on making inferences from tables like

		Witness	
		Blue	Green
Cab	Blue	120	30
	Green	170	680

Sampling of 1000 Cabs

rather than hitting students with a formalistic law like Bayes Theorem. Would they recognize such tasks in judgmental situations?

b) Can instruction be tailored in such a way that the effects of heuristics such as availability, representativeness and anchoring can be diminished in those instances where they mislead the estimator? Can the study (SHAUGHNESSY 1977), which demonstrated some success in overcoming reliance upon such heuristics by implementing an activity-based course in probability and statistics, be replicated? Can instruction such as this maintain its effects over time, or are such results just an "end of course" phenomenon?

2. Causal schemas (KAHNEMAN and TVERSKY 1980) and deterministic reasoning (FISCHBEIN 1976; WELL et. al. 1982; GREEN 1982) have frequently been cited as reasons for the inability of humans to make adequate inferences under uncertainty. Is reliance upon causal schemas or deterministic reasoning a western cultural phenomenon, or can similar results be obtained in cross-cultural studies? Is the causal-schema phenomenon an artifact of the heavy emphasis upon deterministic models in our western culture? If so, can instruction which emphasizes model building and which elucidates the differences between statistical and deterministic models make a difference in people's inferential ability?

3. Most of the research on heuristics and biases has been done with individual subjects. Are groups subject to the same heuristics and biases as individuals? If a tempering effect on probabilistic misconceptions occurs in group processes, should we perhaps teach probability and statistics and judgment and decision making in small groups? A program called Guided Design (WALES 1976) teaches decision making in small groups in introductory engineering courses at West Virginia University.

4. Are professional statisticians subject to these same heuristics and biases, or do they approach judgment and decision-making tasks from a normative point of view? KAHNEMAN and TVERSKY (1971) reported misconceptions among psychologists trained in statistics. For most psychologists, statistics is a tool. For professional statisticians however, statistics is their life's blood, a work of art. It would be worthwhile investigating their responses on some of the judgment and decision-making tasks reported in the literature.

5. The role of the task itself and its influence upon the judgmental process warrants further investigation. Changes in the causal attribution value of variables, or changes in the emotional impact of data, seem to affect

those factors to which people attend when making judgments or decisions. Can changes in the task optimize, or at least improve, inference? If so, what are the implications of this for the way we present stochastic problems to our students or to potential judges?

6. Clinical interviews should be carried out with subjects to attempt to gather validating information for the theories we offer as reasons for their inferential behavior. WELL et. al. (1982) has warned about the dangers of overreliance on summary numerical data for theory development.

In summary, much interdisciplinary investigation needs to take place to assess the state of human statistical illiteracy. Are we forever doomed to be poor decision makers, or can we improve our lot? The (perhaps biased) point of view of a teacher is that many, if not all, of the misconceptions of probability can be addressed cognitively. (This teacher is not yet willing to admit that we are forever doomed to be inept at inference tasks.)

It would be interesting to have a psychologist and a mathematics educator team-teach an introductory course in probability and statistics to bring both sides of the coin to bear upon our students. The general public is not yet sufficiently aware of the importance of probability and statistics, or of the prevalence of judgmental tasks in decision making, or of their interrelationship. Teachers need to become aware of the implications of judgment and decision-making research for the teaching of probability and statistics. We cannot wait for some vivid, concrete event (NISBETT et. al. 1976) to occur which will convince everyone of the importance of probability and statistics when making decisions. Nor can we continue to tolerate the trial-and-error process of human decision making in risky situations (SLOVIC et. al. 1976). We must begin to spread the word ourselves through cooperative efforts.

References

- Abelson, R. P.: 1976. Script processing in attitude formation and decision making. In J. S. Carroll and J. W. Payne (Eds.), *Cognitive and Social Behavior*. Hillsdale: Lawrence Erlbaum.
- Ajzen, Icek: 1977. Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, 35, 5, 303-314.
- Bar-Hillel, M.: 1980. The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211-233.

- Bar-Hillel, M.: 1982. Studies of representativeness. In D. Kahneman, P. Slovic, and A. Tversky (Eds.) *Judgment and Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Bar-Hillel, M. and Falk, R.: 1982. Some teasers concerning conditional probabilities. *Cognition*, 11, 109-122.
- Benz, H. J.: 1982. *Stochastics-Teaching based on common sense*. Osnabrücker Schriften zur Mathematik, Universität Osnabrück.
- Buckley, W. F.: 1982. Column in the New York Daily News, July 27.
- Doumouchel, W. H. and Krantz, D. H.: 1979. Statistics notes. Unpublished manuscript, University of Michigan.
- Falk, R.: 1980. Probabilistic reasoning as an extension of common sense thinking. Presented at the International Conference on Mathematical Education, ICME IV, Berkeley, CA.
- Falk, R.: 1981a. Perception of Randomness. From the proceedings of the fifth international conference for the psychology of mathematics education, Grenoble, France, 222-229.
- Falk, R.: 1981b. On Coincidences. *The Skeptical Inquirer*, 6, 2, 18-31.
- Falk, R. and Bar-Hillel, M.: 1980. Magic possibilities of the weighted average. *Mathematics Magazine*, 106-107.
- Fischbein, E.: 1975. *The intuitive sources of probabilistic thinking in children*. Dordrecht-Boston: D. Reidel Publishing Company.
- Fischbein, E. Probabilistic thinking in children and adolescents. In: IDM (Hrsg.) *Forschung zum Prozess des Mathematiklernens*, Reihe: Materialien und Studien, Band 2, 23-42.
- Fischbein, E.: 1982. Can the teaching of probability improve probabilistic intuitions? From the proceedings of the First International Conference on the Teaching of Statistics, ICOTS I, Sheffield, England.
- Fischbein, E., Pampu, J., and Manzat, J.: 1970a. Companion of ratios and the chance concept in children. *Child Development*, 41, 377-389.
- Fischbein, E., Pampu, J., and Manzat, J.: 1970b. Effects of age and instruction on combinatory ability in children. *British Journal of Educational Psychology*, 40, 261-270.
- Fischhoff, B., Slovic, P., and Lichtenstein, S.: 1979. Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, 23, 339-359.
- Green, D. R. Probability concepts in 11-16 year old pupils. Report from the Center for Advancement of Mathematical Education in Technology, Loughborough University of Technology.
- Holmes, P. and Turner, D.: 1981. Teaching Statistics to 11-16 year olds. In

- A. Shulte (Ed.), *Teaching Statistics and Probability*. The Yearbook of the National Council of Teachers of Mathematics.
- Huff, D.: 1954. *How to Lie with Statistics*. London: W. W. Norton.
- Kahneman, D., and Tversky, A.: 1972. Subjective probability: a judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kahneman, D., and Tversky, A.: 1973a. Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Kahneman, D., and Tversky, A.: 1974. Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.
- Kahneman, D., and Tversky, A.: 1973. On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kahneman, D., Slovic, P., and Tversky, A.: 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Kuhn, T. S.: 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- La Brecque, M.: 1980. On reaching sounder judgments: strategies and snares. *Psychology Today*, June, 33-42.
- Lyon, D., and Slovic, P.: 1976. Dominance of accuracy information and neglect of base-rates in probability estimation. *Acta Psychologica*, 40, 287-298.
- Mathematics Resource Project. *Statistics and Information Organization*. Palo Alto: Creative Publications, 1977.
- Mosteller, F.: 1962. *Fifty Challenging Problems in Probability with Solutions*. Reading: Addison-Wesley.
- Mosteller, F., Kruskal, W. H., Link, R. F., Peiters, R. S., and Rising, G. R.: 1973. *Statistics by Example*. Reading: Addison-Wesley.
- Newell, A. and Simon, H.: 1972. *Human Problem Solving*. New Jersey: Prentice Hall.
- Nisbett, R. E. and Borgida, E.: 1975. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32, 932-943.
- Nisbett, R. E., Borgida, E., Crandall, R., and Reed, H.: 1976. Popular induction: Information is not necessarily informative. In J. S. Carroll and J. Payne (Eds.), *Cognition and Social Behavior*. Hillsdale: Lawrence Erlbaum.
- Nisbett, R. E. and Bellows, N.: 1977. Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, 35, 613-624.
- Nisbett, R. E. and Wilson, T. D.: 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Reviews*, 84, 231-259.

- Nisbett, R. E. and Ross, L.: 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice Hall.
- Rade, L.: 1977. *Take a Chance with Your Calculator: Probability Problems for Programmable Calculators*. Forest Grove, Oregon: Dilithium Press.
- Reinhardt, H. E.: 1981. Some Statistical Paradoxes. In A. Shulte (Ed.), *Teaching Statistics and Probability*, the 1981 National Council of Teachers of Mathematics Yearbook, 100-109.
- Ross, M. and Sicoly, F.: 1982. Egocentric biases in availability and attribution. In Kahneman, D., Slovic, P., and Tversky, A. (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Schoenfeld, A.: 1979. Explicit heuristics training as a variable in problem solving performance. *Journal of Research in Mathematics Education*, 10, 3.
- Scholz, R.: 1981. *Stochastische Problemaufgaben: Analysen aus Didaktischer und Psychologischer Perspektive*. Institut für Didaktik der Mathematik der Universität Bielefeld, Materialien und Studien, Band 23.
- Shaughnessy, J. M.: 1977. Misconceptions of probability: an experiment with a small-group, activity-based model building approach to introductory probability. *Educational Studies in Mathematics*, 8, 295-316.
- Shaughnessy, J. M.: 1981. Misconceptions of probability: from systematic errors to systematic experiments and decisions. In A. Shulte (Ed.), *Teaching Statistics and Probability*, the 1981 National Council of Teachers of Mathematics Yearbook, 90-100.
- Shaughnessy, J. M.: 1982. Misconceptions of probability, systematic and otherwise: Teaching probability and statistics so as to overcome some misconceptions. Presented at the First International Conference on the Teaching of Statistics, Sheffield, England.
- Shulte, A.: 1981. *Teaching Statistics and Probability*. The 1981 yearbook of the National Council of Teachers of Mathematics.
- Slovic, P., Fischhoff, B., and Lichtenstein, S.: 1976. Cognitive processes and societal risk taking. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and Social Behavior*. Hillsdale: Lawrence Erlbaum.
- Slovic, P., Fischhoff, B., and Lichtenstein, S.: 1977. Behavioral decision theory. *Annual Review of Psychology*, 28, 1-39.
- Slovic, P., Fischhoff, B., and Lichtenstein, S.: 1980. Risky assumptions. *Psychology Today*, June, 14, 44-48.
- Slovic, P.: 1982. Toward understanding and improving decisions. In W. Howell (Ed.), *Human Performance and Productivity*. Hillsdale: Lawrence Erlbaum.
- Taylor, S. E.: 1982. The availability bias in social perception and interaction. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

"Trial by Mathematics": 1968. *Time*, 26, April, p. 41.

Tversky, A., and Kahneman, D.: 1971. Belief in the Law of Small Numbers. *Psychological Bulletin*, 76, 105-110.

Tversky, A., and Kahneman, D. Causal Schemata in Judgment under Uncertainty. In M. Fischbein (Ed.), *Progress in Social Psychology*.

Wagner, C. H.: 1981. Paradoxes in Sampling. In *Teaching Statistics and Probability*, the 1981 National Council of Teachers of Mathematics Yearbook, Ed. by A. Shulte, 162-167.

Well, A. D., Pollatsek, A., and Konold, C. Probability Estimation and the Use and Neglect of Base-Rate Information. Draft copy of a paper presented at the research pre-session of the 1982 National Council of Teachers of Mathematics meeting in Toronto. Under revision at the University of Massachusetts, Amherst.

(MIS-) INTERPRETATION OF STOCHASTIC MODELS

GEORG SCHRAGE

*Universität Dortmund
Institut für Didaktik der Mathematik
F.R. Germany*

I am not an expert in the theory of decision making, rather I am concerned with mathematics education and teacher training. Within the frame of this task one of my special interests is teaching statistics. Statistics - to use a definition of W. A. Wallis and H. W. Roberts - is a body of methods for making wise decisions in the face of uncertainty. And that is the subject of our conference.

In german schools, teaching statistics is part of mathematics education. To be precise, there are not special courses in statistics but statistical applications are involved in a stochastic curriculum that is part of the mathematics curriculum from primary school up to college level. That may sound very good and certainly there has been an important development during the past decade. But reality of statistics education in german schools is still far from being satisfactory and I doubt that it brings about an improved decision behavior of students. Some related test results are later presented.

In my judgement the integration of stochastic elements has succeeded best in primary schools, where children gather experiences with random behavior by performing and observation of real experiments. The aim at this primary level should be to establish fundamental concepts and intuitive feeling for the regularities of random outcome.

At the level of age ten to sixteen - the german Sekundarstufe I - these experiences should be ordered and made explicit. Quantitative aspects are emphasized. Probabilities of events that are connected with simple Laplace-experiments are calculated. Statistics in these classes should be mainly descriptive statistics. There are very stimulating appropriate materials for classroom investigations. Especially Prof. Winter from the University in Aachen made a lot of exciting proposals how statistics can be integrated in mathematics education at this level. But reality - as I see it - is that most pupils at this age are never confronted at school with any statistical problem. The main reasons for this lack are teaching traditions, inflexibility of the german school systems and insufficient prepared teachers.

At the upper secondary level - the german Sekundarstufe II - students have to choose their courses. Nowadays many schools offer stochastic courses and I estimate that between 10 and 20 percent of pupils participate. But these courses are often unsatisfactory because statistical inference is reduced to mathematics. Students get knowledge of the most common distribution functions, they learn to estimate parameters, to make tests depending on a fixed level of significance and perhaps to calculate intervals of confidence. Examples are most of such a kind that there is no doubt about the corresponding mathematical model, so that standardized methods of inference can be applied. In most german text books the character of statistics as a methodology for making decisions, taking into consideration the consequences of possible errors, is completely hidden or not existent. So statistics becomes a branch of mathematics like calculus, geometry or algebra. But mathematics is only one face of statistics. Usually the real difficulties of a statistical investigation have to be solved before the process of decision making can be reduced to a mathematical problem. Therefore I think that statistics education must include the nonmathematical aspects of statistical projects: The precise formulation of the problem, the process of modeling, the discussion of reasonable alternative decisions and their possible consequences, questions referring to risk and personal responsibility, the process of sampling and the interpretation or even the application of the results

received by statistical inference. This can be done best by carrying out some real projects. At the symposium on stochastics in education, that took place 1980 in Klagenfurt, I presented a course, emphasizing decision-theoretical aspects and developing fundamental ideas of statistical inference by means of some realistic projects that are suitable to be realized at school.

It is a matter of fact, and I will prove it by some examples, that even people with some mathematical or statistical background may use formal methods in misleading ways. To overcome this situation, educators must get more knowledge on defective concepts and strategies. What we would need is an elaborated phenomenology of frequent errors and misleading strategies. There are already many systematical investigations on stochastic conceptions and misconceptions. Especially I refer to the work of Mr. Shaughnessy and his stimulating article "Misconceptions of Probability: From Systematic Errors to Systematic Experiments and Decisions" in the 1981 Yearbook on Teaching Statistics and Probability of the NCTM. But we do not know enough of the reasons for misconceptions and we do not know enough of the influence of education on misconceptions. At the Sheffield conference on teaching statistics in August of this year, Prof. Fischbein reported on some tests and pointed out that stochastics education can even be responsible for misconceptions and loss of good intuition. At the same conference Mr. Green of the Loughborough University of Technology reported on a project to investigate stochastic conceptions of about 3000 children from 11 to 16. I will present you two of the test items.

1. Which of the following results is more likely?

- a) Getting 7 or more boys out of the first 10 babies born in a new hospital.
 - b) Getting 70 or more boys out of the first 100 babies born in a new hospital.
- (A) They are equally likely
 - (B) 7 or more out of 10 is more likely
 - (C) 70 or more out of 100 is more likely
 - (D) No one can say.

I think this is a very good question to test students understanding of long run stability of relative frequencies.

The result:

- (A) - 25 %
- (B) - 8 %
- (C) - 5 %
- (D) - 61 %

I asked the same question to 153 of my teacher students at the University of Dortmund, omitting the possibility (D).

The result:

- (A) - 60 %
- (B) - 14 %
- (C) - 15 %
- no answer - 10 %

The result is still worse than it seems, because a great part of the students who answered correctly gave wrong arguments for their answer. Nobody referred to long run stability of frequencies. From 17 students, who had participated at a stochastic course at school, only one gave the correct answer. Many students calculated the probability for both events to be $\frac{7}{10}$. By the way, Mr. Shaughnessy in the above mentioned article reports on the same question and similar results.

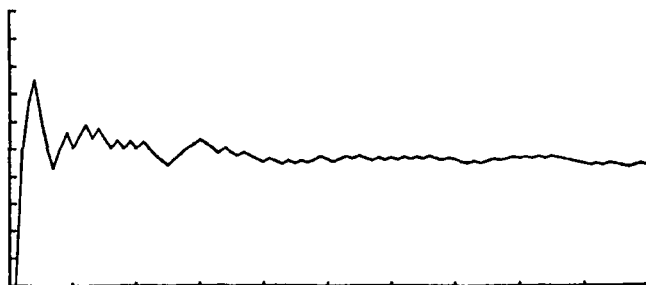
Some weeks ago I had a disillusioning experience. At the beginning of my lectures on stochastics I talked in detail on the empirical background of probability theory. I spoke on long run stability of frequencies, I showed the graph of a real experiment's outcome as below and I put at the black board the empirical law of large numbers.

Then I asked my students again the above question. The result was approximately the same as in the test. My test also contained items to prove students understanding of independence. The result was not better than the result of the just mentioned item.

```

0 1 1 1 0 0 0 1 : 0 1 1 0 1 0 0 1 0 1 0 0 0 0 1 1 1 1 0 0 0 1 0 0
1 0 0 0 1 0 0 1 0 1 0 1 1 0 0 1 1 0 1 0 0 1 0 1 0 0 0 0
1 0 1 1 0 1 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 1 1 0

```



2. A bag has in it some white and some black balls. A boy picks out a ball, notes its colour and puts it back. He then shakes the bag. He does this four times. He picks a black ball every time. Then he picks out one more ball.

What a colour do you think he is likely to get?

- (A) Black is more likely again
- (B) Black and white are equally likely
- (C) White is more likely this time.

The answer should be (A).

I think this is a bad question. As a consequence of the absence of any a priori probability one can say absolutely nothing on the probability to take a white or a black ball out of the bag. If I say that it is more likely to get a black one, then that is a pure subjective judgement. The classification as "true" or "false" is not appropriate for this situation. If I had to make a decision, perhaps I would decide for "black". But making a decision and making a statement on the probability of an event are two completely different pairs of shoes. Believing that a hypothesis that has been rejected by a statistical test is unlikely to be true is a very common misinterpretation. This is an error similar to base rate fallacies. In an article on teaching probability theory and statistics I found the following interpretation of a test:

If H_0 is true then the event E is nearly sure
 The outcome of an experiment was $\bar{E}$

So H_0 is nearly impossible

In 1975 I observed a murder trial that took place in Wuppertal - a town about 100 miles away from Bielefeld - where this way of thinking nearly had produced an error of justice. An unknown man had murdered a woman, and there was a defendant who was committed for trial since two years. The law court asked some experts on their opinion about some accusing circumstances. One expert - a member of the Bundeskriminalamt (the German FBI) - had found at the defendant a tiny proof - a millionth part of a gram - of stuff of the same kind as the clothes of the killed woman. The expert now argued in the following way: The probability to find proofs of such stuff at a person who was not in contact with the killed woman is 0,06 percent. (I don't know how he got this number.) So the defendant had contact with the killed woman and must be considered to be the murderer with probability 0,9994. In the same way argued a doctor of medicine who called himself a biostatistical expert. He had analysed a tiny proof of blood, that was used up by the analysis. He calculated a probability of 0,979 that the defendant was guilty. All the world including the tribunal believed that the defendant's guilt was mathematically proved.

In this situation I placed myself at the disposal of the attorney for the defence. In my comment I contradicted the argumentation of the experts by the following example: Some time ago a bank was robbed by a red-haired man. About 1 percent of the German population is red-haired. So, if I was not the robber, the probability for a red-haired robber would have been 1 percent. Consequently the experts would calculate a probability of 99 percent that I was the robber.

This is an example how an erroneous kind of argument can be unmasked by only changing the context. By the way, the defendant was later acquitted, since he was proven to be innocent.

I believe that the most frequent types of misinterpretation of stochastic models are generated by imprecise common use of language. v. Neumann and Morgenstern in their book on game theory described this problem: "There are notions which are quite fundamental for our discussion but the use of which in everyday language is highly ambiguous. The words which describe

them are used sometimes in one sense, sometimes in another and occasionally - worst of all - as if they were synonyms."

We speak of the probability of an event. But an event for itself does not obey any attribute that can be called its probability. The probability of an event is part of a model that describes either frequencies depending on certain experimental conditions, or the subjective degree of belief that may be reflected by the betting behavior of a person. So if we make a model of a stochastic situation, the interpretation of corresponding statements must be very clear. I will give you an example where the relationship between the problem that should be investigated and the stochastic model has been totally confused.

In 1972 a journal for mathematics teachers published an article with the title "Mit welcher Wahrscheinlichkeit gilt der Fermat-Satz?" (What is the probability that Fermat's conjecture is true?) Fermat's conjecture is one of the challenging unsolved problems of number theory. It states that the equation $x^n + y^n = z^n$ has no solution in natural numbers if $n \geq 3$. The author of the mentioned article calculated the probability of the truth of the conjecture as follows:



He considered an interval of natural numbers $\mathbb{N}_k = \{1, 2, \dots, k\}$. Let M_1^k be the subset of numbers that can be represented in the form $x^n + y^n$ for certain n and M_2^k the subset of numbers that can be represented in the form z^n , i.e.,

$$M_1^k = \{a \in \mathbb{N}_k \mid a = x^n + y^n, \quad x, y \in \mathbb{N}\} \quad \text{and}$$

$$M_2^k = \{a \in \mathbb{N}_k \mid a = z^n, \quad z \in \mathbb{N}\}.$$

Fermat's conjecture asserts that $M_2^k \cap M_1^k = \emptyset$ for any k if $n \geq 3$.

The author considered the interval $\mathbb{N}_k$ to be an urn and made a stochastic model for the experiment to take first a subset $A^k \subset \mathbb{N}_k$ such that $\# A^k = \# M_1^k$ - assigning equal probability to each possible outcome - and then a second subset $B^k \subset \mathbb{N}_k$ such that $\# B^k = \# M_2^k$. He calculated the probability $P_k = P(A^k \cap B^k = \emptyset)$. This value was interpreted to be the probability that no counterexample to Fermat's conjecture can be found in the urn $\mathbb{N}_k$.

And the result $\lim_{k \rightarrow \infty} P_k = 1$ was accordingly interpreted as the probability of the truth of Fermat's conjecture.

But there is no justification for this interpretation. The model represents a hypothetical stochastic experiment and the result of the calculation may be interpreted as a statement on frequencies related to this experiment. The truth of Fermat's conjecture however is completely independent of any random experiment and its possible outcome. We cannot calculate any objective probability for Fermat's conjecture. Moreover the above interpretation leads to the absurd consequence that the probability of a counterexample in the urn N_k is decreasing when k is increasing. Nevertheless - and this is the important thing - there was no protest of the readers. Quite on the contrary, an editorial committee including some university mathematicians awarded a prize to this article.

I will come to an end with an example that demonstrates some experts lack of commonsense. German students to get admission for studying medicine must pass through a test. The tests are developed with great expense by a central institute. In 1979 this institute made public some typical items, among them the following:¹⁾

The table below shows the frequencies (in percent) of five symptoms with respect to the maladies 1 to 5.

symptoms	malady				
	1	2	3	4	5
Insomnia	80	90	75	90	95
Headache	50	40	65	30	60
Trembling	60	20	50	80	10
Falls	70	30	30	60	60
Skin rash	10	50	0	50	30

1) *Der Spiegel*, Nr. 14/1979

Mr. Meyer comes to the doctor and complains of insomnia, falls and trembling.

His malady will probably be

1	2	3	4	5
☐	☐	☐	☐	☐

The question was considered of moderate difficulty. The "correct" answer should be 4.

It should be clear that there is not enough information to answer the question. We know nothing on the base rate of the different maladies and nothing on the dependence of symptoms. Moreover, under the little realistic assumptions that all maladies have the same base rate and that all symptoms are mutually independent, both 1 and 4 have the same probability 0,404.

How did the experts of the institute get their answer? They used the following algorithm: For each of the maladies add the probabilities of the patient's symptoms. The illness with maximal corresponding sum will be the most probable one. Why? I do not know. This algorithm has absolutely nothing to do with the problem. For example, it ignores the fact that the absence of certain symptoms gives information on the kind of malady as well as the presence of other symptoms.

Since intelligent students will be misled by such poor test items, I wrote a letter to the people who prepared the test. I exhibited the following example to demonstrate the wrongness of their thinking:

Let us assume that 99 % of people suffering from cancer of lung sometimes get a fit of coughing while the same symptom is observed at 50 % of people who caught a cold.

symptom	malady	
	cancer	cold
coughing	99	50

Now suppose Mr. Meyer gets a fit of cough while he consults to a doctor who learned decision making according to the just mentioned algorithm.

My example did not convince the experts of the test institute. I got an answer including detailed mathematical explanations that I did not understand. But there was a further remarkable argument: "The test item was examined by more than thirty university teachers and practitioners including experts of biomathematics and statistics. None of these experts criticized the question or the kind of solution. Experiences with the test proved that the question is correctly understood and answered by most students."

I tested this statement with my students, and really most of them used the same algorithm as the test engineers.

I think it should be a central aim of statistics education that students recognize such serious abuse of stochastic instruments. But what can we do to reach this aim? One point should be to discuss in the classroom examples of misleading strategies to warn students never to lose the meaning of mathematical operations.

To overcome misconceptions, Mr. Shaughnessy at the Sheffield conference recommended the orientation of stochastics teaching at some guiding principles that I will repeat:

1. Introductory courses should be activity based and experimental,
2. Emphasize judgement and decision making,
3. Sensitize students to the misuse of statistics,
4. Give very stimulating problems.

I agree to this program. It should be realized by statistics teachers in their daily work in the classroom and so it is the task of my colleagues and me to prepare teachers to be capable to accomplish such a program.

References

- Schrage, G.: 1980. Schwierigkeiten mit stochastischer Modellbildung. Zwei Beispiele aus der Praxis. In: JDM 1, 86-101
- Schrage, G.: 1981. Entscheiden und begründen - Leitlinien für den Statistikunterricht. In: W. Dörfler & R. Fischer (eds.), Stochastik im Schulunterricht. Schriftenreihe Didaktik der Mathematik, Universität für Bildungswissenschaften in Klagenfurt, Band 3, Wien
- Shaughnessy, J.M.: 1981. Misconceptions of probability: From systematic errors to systematic experiments and decisions. In: A. Shulte (ed.), Teaching Statistics and Probability. The 1981 National Council of Teachers of Mathematics Yearbook, 90-100
- Winter, H.: 1981. Zur Beschreibenden Statistik in der Sekundarstufe I - Rechtfertigungsgründe und Möglichkeiten zur Integration der Stochastik in den Mathematikunterricht. In: W. Dörfler & R. Fischer (eds.), Stochastik im Schulunterricht. Schriftenreihe Didaktik der Mathematik, Universität für Bildungswissenschaften in Klagenfurt, Band 3, Wien

This Page Intentionally Left Blank

**– RANDOMNESS AND STOCHASTIC INDEPENDENCE –
ON THE RELATIONSHIP BETWEEN INTUITIVE MOTION
MATHEMATICAL DEFINITION**

GERD v. HARTEN

HEINZ STEINBRING

*Universität Bielefeld
Institut für Didaktik der Mathematik
F.R. Germany*

1. Introduction

In stochastics education, one of the most important questions is how students handle the concepts of chance and of randomness. It refers to two aspects: first, to the precepts and intuitions of probability, and second, to the mathematical calculi and techniques used to adapt and to develop these intuitions further.

The concept of chance or of randomness is central for stochastics education for the very reason that this concept, as opposed to other mathematical theories and concepts, is the specific feature of probability theory.

Stochastics independence has emerged, during the historical development of probability theory, as one central mathematical technique of assigning a mathematically precise meaning to randomness. Nevertheless, it should be mentioned that this does not solve the problem of intuition and of calculus of chance. Independence itself reflects the same problematical relationship: "What is really involved is a definition of independence and a belief (borne out by experience and experiment, to be sure) that the definition is applicable to a particular situation. There is, thus, independence in a

vague and intuitive sense, and there is 'independence' in the narrow but well-defined sense that the rule of multiplication of probability is applicable. It was the vague and intuitive notion that provided for a long time the main motivation and driving force behind probability theory." (KAC, 1959, p.10)

Chance and randomness is a common fundamental assumption both of stochastic education and a great part of psychological research in decision making under uncertainty.

In the following paper, we shall analyze some aspects of this topic from an educational perspective, and shall try to find out some similarities or differences between these two approaches; i.e. we shall see what the specific educational requirements for investigating pupils' intuitions and pre-concepts are.

2. How can randomness be tested?

One central problem in psychological research about stochastic thinking is how to evaluate individual judgments of randomness. Usually, various binary sequences are presented to the student who is asked to estimate which of these sequences is random. The students will mostly only be able to give intuitive judgments as they lack detailed knowledge in probability theory. The adequacy of these answers is discussed as compared to that of answers given by an expert or obtained by formal calculus. The latter might provide the correct answer, i.e. which sequence is random, and which is not.

David Green made a survey of probability concepts of 3000 pupils aged 11-16 years. In his work he asked the following questions:

A teacher asked Clare and Susan each to toss a coin a large number of times and to record every time whether the coin landed Heads or Tails. For each 'Heads' a 1 is recorded and for each 'Tails' a 0 is recorded.

Here are the two sets of results:

Clare: 01011001100101011011010001110001101101010110010001
 01010011100110101100101100101100100101110110011011
 01010010110010101100010011010110011101110101100011
 Susan: 10011101111010011100100111001000111011111101010101
 1110000001000101001000001000110001010000000011001
 01000001111100001101010010010011111101001100011000

Now one girl did it properly, by tossing the coin. The other girl cheated and just made it up.

- (a) Which girl cheated? Answer
- (b) How can you tell? Answer

The results are shown in the following table:

Year	Clare	Susan	'Cannot tell'	Omit
1	38	53	1	7
2	34	56	2	7
3	30	59	4	7
4	40	46	3	9
5	35	49	5	9
All	35	53	3	8

Posing such or similar problems serves to thematize the relationship between individual intuitions and objective, or theoretically correct answers. Research into intuitive precepts of randomness directly refers to naive ideas about randomness and chance in the sense of "undetermined", "no repetitive or regular structure", "no possibility of predicting further outcomes", etc.

In the attempt to precisely define the concept of randomness, the so-called theory of computational complexity adopts a quite similar view of chance (Cf. FINE 1973, p.118 ff). Roughly speaking, a finite binary sequence is called random if the smallest algorithm capable of specifying it to a computer has about the same number of information bits as the sequence itself

(CHAITIN 1975, p.48). This is just what would be naively expected, as a sequence for which an algorithm (or a specific generating law) can be constructed which is "smaller" than this sequence, condensing it, should not be random in the sense of this definition. It is clear that computer-generated random numbers, according to this definition, are not really random but rather pseudo-random.

It is seen that this approach to randomness within the theory of computational complexity corresponds to some aspects of preconcepts and intuitions of chance and randomness.

The attempt at formalizing such intuitions in a theoretical frame permits to draw some interesting epistemological (and mathematical) conclusions. These conclusions will shed some light onto the difficult relation between intuition and theoretical techniques. CHAITIN (1976) has shown that this definition of randomness corresponding to intuition cannot be proved in the following sense: There are sequences of zeros and ones, of which we cannot tell whether they are random, or not. In one word, this is unprovable.

The conclude that our definition of randomness is generally unacceptable would mean emptying the baby out with the bath. We should like to propose the following distinction: on the one hand, the unprovability of randomness implies that we cannot absolutely or universally decide whether a sequence is random or not; on the other hand, however, we have to be content with a relative decision about randomness. In quite the same sense as it is used in statistical practice, a sequence is pragmatically called random if it satisfies all the known or relevant statistical tests. Relative to the body of test techniques and instruments, a sequence will be more or less random. There is no randomness in an absolute sense. This distinction explains why a computer-generated random sequence - which is not random in an absolute sense - is random in a relative sense, and why such sequences will do in statistical simulation.

One might draw on analogy to the principle of relativity stated in the following form: "... if a space ship is drifting along at a uniform speed, all experiments performed in the space ship and all phenomena in the space ship

will appear the same as if the ship were not moving ..." (FEYNMAN, 1970, p. 15-1). This means, the tools i.e. experiments for detecting motion become crucial.

How could this analysis contribute fruitfully to the problem of studying and developing the concept of randomness in the understanding of pupils?

First of all, we should keep an eye on the close interrelation between intuitive ideas about randomness and the body of test techniques available to deal with this concept. In particular this means that both ideas and intuitions, as well as mathematical techniques and rules, have to be taken into account at every level of understanding, no matter how provisional they may be.

Intuition and calculus of chance are not contradictory, but calculus should be used to further develop and clarify intuitions which, however, will not be definite.

For stochastics education - and perhaps in some aspects for psychological studies as well - we have to draw the following conclusions: In order to unfold stochastical thinking, it is not sufficient to confront individuals with single random situations. Besides, it is required to be aware of the means available to describe and analyze this situation. In school mathematics, we have to construct not single tasks, but rather connected sequences of tasks in which the degree of difficulty, the scope of techniques and mathematical concepts required are varied and which will show the central meaning of different approaches and variable modellings for the interpretation of real random events.

3. Independence - intuitive notion and mathematical calculus

A typical problem in the stochastics section of school textbooks reads as follows: "Two dice are thrown. What is the probability that the sum of scores is greater or equal 10 if it is known that one die shows a 4?" Pupils are expected to find the following solution: "There are 36 possible outcomes of this experiment. If it is known that one of the dice shows a 4, then there

are only 11 outcomes left. Among these, only 2 will yield a sum of scores greater or equal 10. Hence, the probability is $2/11$."

H. Dinges reports that he presented this problem to some of his colleagues of the mathematics department with the result that they either asked him to describe the situation more precisely, or gave the solution of $1/6$. This answer is due to the following interpretation of this situation:

Two dice have been thrown, and an observer has glimpsed that one of them shows a 4. With which probability may the observer expect a sum of scores higher than 10?

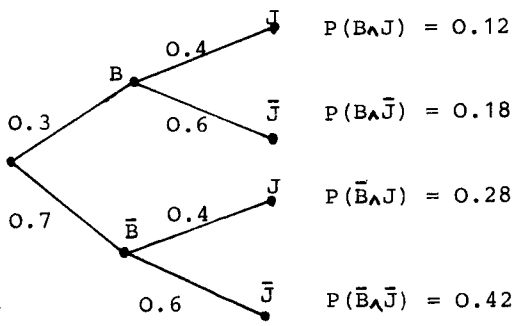
It is seen that the textbook problem is open to representation by different models. While the textbook authors appear to have presented the situation as a paradigm of applying conditional probability, experts confronted with the problem tried to imagine a real situation.

Let us now consider a concept which is at the very core of probability theory and which ought to be of prime importance for the school curriculum as well - the concept of independence which is the central instrument to handle randomness in school mathematics. The definition of stochastic independence is usually given for two events as follows:

Two events A, B are independent if $P(A \cap B) = P(A) \cdot P(B)$.

To use this concept for solving simple problems, we must have some intuitive notion of independence which tells us whether or not a chosen model is adequate for a certain situation. In school, such models are used e.g. for two-stage experiments as the throwing of two dice. In more sophisticated situations, the model character of independence becomes more obvious. Let us take a look at the following textbook example:

"Bill and Joe are two frequent guests in a pub. The innkeeper knows by experience that Bill will take a morning pint with a probability of 30 percent, while Joe's probability is 40 percent. The two often quarrel. Hence, the innkeeper prefers either Bill or Joe coming alone to take a morning pint. What is the probability for this event?"



The solution of this example is based on the hypothesis of independence. Independence is used as a tool to construct a probability space for the compound event; without this hypothesis, there is no solution. We can consider this use to be the constructive aspect of independence.

Let us consider another problem: "A coin is tossed and two observers X, Y are watching the outcome E. The two observers make no collusion. Are they 'independent'?" Let H = heads, T = tails, $P(E=T) = p$

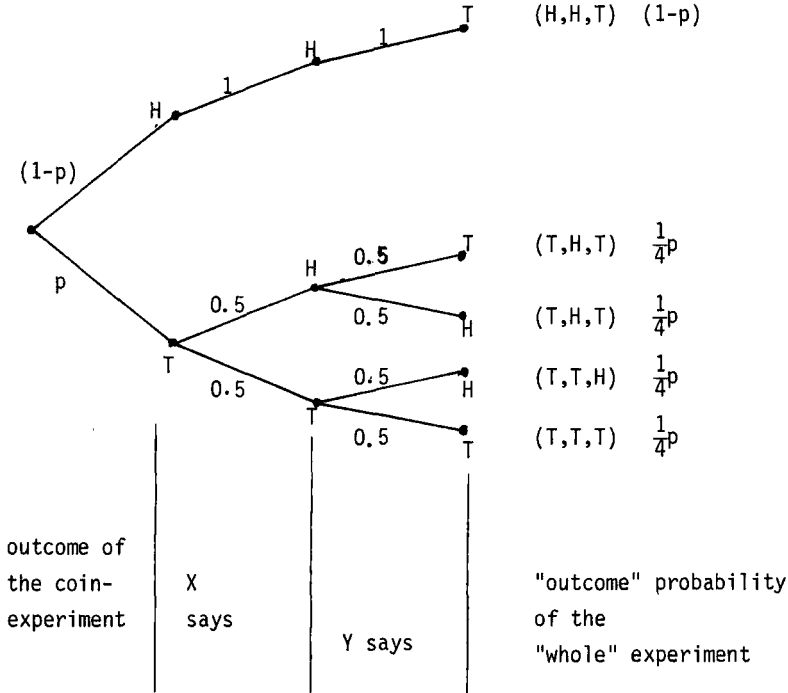
Statement	true outcome	
	H	T
'H'	$(1-p)$	$\frac{1}{2} \cdot p$
'T'	0	$\frac{1}{2} \cdot p$

probabilities

Statement	true outcome	
	H	T
'H'	0	$\frac{1}{2} \cdot p$
'T'	$(1-p)$	$\frac{1}{2} \cdot p$

probabilities

Now we can easily compute that the events "X tells the truth" and "Y tells the truth" are not independent (only in the exceptional cases of $p = 0$ or 1).



$$P(X \text{ says the truth}) = (1-p) + \frac{1}{2}p = 1 - \frac{1}{2}p$$

$$P(Y \text{ says the truth}) = \frac{1}{4}p + \frac{1}{4}p = \frac{1}{2}p$$

$$P(X \text{ says the truth and } Y \text{ says the truth}) = \frac{1}{4}p$$

$$P(X \text{ says the truth}) \cdot P(Y \text{ says the truth}) = \frac{1}{2}p(1 - \frac{1}{2}p) = p(\frac{1}{2} - \frac{1}{4}p)$$

This result is evident as the probability of these events is linked to the experiment's outcome. The events are only conditionally independent, with respect to A, the outcome of the experiment. Thus we are led to the following

well-known notion:

Two events A, B are conditionally independent given an event C iff

$$P(A \cap B | C) = P(A | C) \cdot (P(B | C))$$

In this example, we have used the formal mathematical concept of independence to discover that some events are not independent, although the observers have not been in collusion, or might even have been in different rooms. (Condorcet, Laplace, Poisson, Cournot, Boole have drawn some conclusions with regard to judgments founded on the assumption that if two judges are not in collusion the events of their telling the truth will be stochastically independent.)

This example shows another aspect of independence, which we will call the exploratory aspect of independence. This reflects the fact that independence is not only used to describe but to discover independence. Let us come to a last example before we might draw some educational propositions: "In an urn, there are 50 balls: 49 black and 1 white. Two persons alternately draw without replacement. He who draws the white ball first has won. Does the first to begin or the second have the better chance to win?"

To solve this problem one can assume that both players continue to draw until the urn is empty. Now all possible sequences of black and white balls have the same probability. Therefore getting a white ball at an even draw or an uneven is equally likely, provided the number of draws is even.

This condition, we have noticed in this special case, is generally known as the concept of exchangeability defined usually in the following way: A family $\{X_i\}$ of random variables is called exchangeable, if for every finite collection and every permutation σ :

$$F_{X_1, \dots, X_n} = F_{X_\sigma, \dots, X_{\sigma(n)}}$$

where F denotes the distribution function. Thus from such a simple problem, pupils might gain first insights in a fundamental concept as exchangeability which plays a central role in subjective probability theory.

4. *Educational proposition*

All the examples we have given served to illustrate the following point: to apply the mathematical concepts of independence, we must possess an intuitive notion; where the latter may be inadequate, mathematical concepts as independence for example will serve as a tool for investigation. We are called to construct a more or less adequate model of a real situation.

For this purpose, we want to recall the difference between a theory and its application. In the 18th century, such a difference did not exist. Probability was directly linked to reason; probability theory was intended to describe how a reasonable being would act. If calculus predicted an unreasonable result, calculus was to blame.

During the subsequent development of probability theory, it was impossible to maintain this kind of direct link between a theory and its application, between concept and content, between the theoretical term and the object. In the end, Kolmogorov's system of axioms was to describe probability only implicitly as a function fulfilling certain relations.

From an educational point of view at least, this axiomatic characterization is not very satisfactory, as it is impossible in school to teach but the use of a concept without explaining its content.

As the above examples have shown, some insight into the meaning of say, independence is required before this concept can be used. On the other hand, the theoretical term of stochastic independence cannot be described by just telling something about its meaning.

A first educational conclusion suggested might be the following: Pupils must learn to construct models; this can be achieved by confronting them with open problems etc. As we have seen, the problems taken from textbooks are usually understood to have a rather clear-cut solution. To the pupil, it is obvious that to solve this problem he/she must apply, for example, the formula of independence, conditional probability, etc. There are but two options: either the pupil knows these concepts and is able to solve the prob-

lem, or he/she will fail. Now this could be easily overcome: just provide the pupils with open problems, where different models are possible.

We do not consider this solution very sound, however. It is not without reason that the context of problems prescribed in the school curriculum is narrow. First, pupils are not researchers (a fact sometimes overlooked by those propagating problem-solving lessons), and finding new theorems is not the objective of instruction. Second, whether or not a problem is solved, or at least a solution adequate in some sense is attained, will depend of the tools used (by this, we not only mean tools in the material sense, but also scientific concepts). Hence, knowledge cannot be confined to procedural skills.

In our view, a possible way out of this apparent dilemma is to treat in school not only singular problems within a strictly fixed context, but to consider, with respect to a certain feature of a theoretical concept or rule, systems of problems which permit to establish a more open context and which reflect, and use as a driving force, the inherent difference between theoretical terms and their applications.

References

- Athen, H. & Griesel, H. (eds.): *Mathematik heute* (Bd.10). Hannover u.a. 1976
- Chaitin, G.J.: *Randomness and mathematical proof*. *Scientific American*, 1975, 232, 47-52.
- Dinges, H.: *Schwierigkeiten mit der Bayesschen Regel*, in: *Math.Phys.Sem.*, Bd. XXI, 1978, 113-156.
- Feynman, R.P. u.a.: *Lectures on Physics*, Vol. I. Reading (Mass.) u.a. 1970.
- Fine, T.L.: *Theories of Probability*. New York 1973.
- Green, D.: *Probability Concepts in 11-16 Year old Pupils*. Manuscript, Loughborough, May 1982.
- Kac, M.: *Statistical Independence in Probability, Analysis and Number Theory*. London u.a. 1964.
- Keynes, J.M.: *Über Wahrscheinlichkeit*. Leipzig 1926.

This Page Intentionally Left Blank

DECISION-THEORETIC PARADOXES

GEORG PFLUG

*Universität Giessen
Fachbereich Mathematik
F. R. Germany*

The history of inductive statistics up to the mathematical foundation of decision theory by Abraham WALD (around the middle of the 20th century) was characterized by the fact that intuitive principles were considered to be of equal importance as mathematical deduction. Such "intuitive principles" were for instance: the optimality of the maximum likelihood (ML-) estimator, the arithmetic mean as best estimator for location, the consistency of all usual estimators.

But soon several examples were found showing that some intuitive concepts of optimality of statistical procedures needed to be revised. Such examples - some of them are presented in this paper - are at the same time counterexamples and paradoxes. The qualification of an example to be astonishing, contra-intuitive or paradox depends on the general state of knowledge and is therefore relative. It can happen that no example amazes today an expert in estimation theory even if he notices this example for the first time. This relative validity however is common to all paradoxes, a fact which is e.g. described by SCHOLZ (1981) for the well known paradox of ZENON.

Nevertheless some of the presented examples shake the theory of statistical decisions to its very foundations. This effect is also common to all paradoxes (Compare the effect of RUSSEL's paradox to the development of mathematical logic). Charles Stein describes in his paper on inadmissibility how sceptical he was about the correctness of his own results - he could not believe such a contra-intuitive situation. And Lindley's Paradox indicates that the Bayesian standpoint can lead to contradictions with classical statistics.

Why does it happen that contra-intuitive situations occur? An partial answer to that question is given by the *theory of restricted rationality*. This theory explains the phenomena by the limited capacitive and/or operative abilities of the human thinking. Although such an explanation is sufficient for the understanding of simple fallacies for children it is not able to explain fallacies of experts. Those fallacies can better be explained by the argument that the paradox examples come mostly from very extreme case. Our intuition which was trained with rather normal cases will be mislead by very unusual examples. It is not surprising that Edgeworth's counterexample (Example 1) makes use of a unusual density and that Stein's Paradox (Example 2) is valid only for $n \geq 3$. So we conjecture that if a statistical situation is considered in an intuitive manner, simple and beautiful examples are kept in mind at the same time, a fact which can lead to the conclusion that there are more analogies than actually are.

Before going into the presentation of the examples some basic facts concerning WALD's statistical decision theory are discussed. Let $f(x, \theta)$ be a family of probability densities which depend on an unknown parameter θ . A function $g(\theta)$, mapping the parameter onto a point on the real axis is to be estimated by an estimator $T(x)$ which depends only on the observations. The "costs" for the discrepancy between the estimated value $T(x)$ and the true value $g(\theta)$ are measured by the *loss function*

$$L(T(x), g(\theta))$$

the mean loss, i.e. the expectation of the loss function with respect to $f(x, \theta)$ is called *risk function*

$$R(T, \theta) := E_{\theta}(L(T(\cdot), g(\theta)))$$

The estimator T_1 is called better than T_2 if

$$R(T_1, \theta) \stackrel{\leq}{\neq} R(T_2, \theta) \quad \text{for all } \theta$$

An estimator is called admissible, if there is no better estimator.

Example 1. An iron trading company has bought shock-absorber tubes at a very low price from a factory which was in bankruptcy. The tubes were without any internal cylinder. Unfortunately the diameter of these tubes proved to be very different. (Perhaps that was the reason why the producing company went into a

bankruptcy?) The distribution of the diameters is unimodal and symmetric. A sample of size n is drawn to determine the mean diameter according to which the appropriate cylinders should be produced. It is possible that the very first observation is a better estimate than the arithmetic mean of repeated measurements.

Explanation. This is Edgeworth's Paradox. The fact that one single observation may be better than the arithmetic mean of repeated observations stands in evident contradiction to the intuitive principle of the optimality of the arithmetic mean which was generally accepted in the 19th century.

The original formulation of this example by EDGEWORTH makes use of the density family

$$f(x, \theta) = \frac{3}{2} \frac{1}{(1+|x-\theta|)^4}$$

and a risk function chosen as follows: Since both parts fit together only if some tolerance limit ϵ is not exceeded it makes sense to set

$$R(T, \theta) = P\{|T - \theta| \leq \epsilon\}$$

By some simple calculations one really finds that for the density defined above and for $\epsilon < 0.1$

$$T_1 = X_1 \text{ is a better estimator than } T_2 = \frac{X_1 + X_2}{2}$$

i.e. T_1 has smaller risk. Analogously, for any given n , it is possible to construct examples that

$$T_1 = X_1 \text{ is a better estimator than } T_2 = \frac{X_1 + X_2 + \dots + X_n}{n}$$

The explanation for this phenomenon is simple. Edgeworth's density has a sharp peak at the center of symmetry. Therefore it is more probable that X_1 lies in the considered confidence interval than $\frac{X_1 + X_2}{2}$, the density of which has not a similar peak. However, this fact is hardly seen intuitively.

Sometimes it is argued that the example shows that in some situations the arithmetic mean of independent observations makes no sense at all. This argument is in the author's opinion not convincing. A different example which goes back to the logician Venn is better suited for this purpose:

Two sailors from a ship are sent to steal into a fort and take the calibre of a cannon. We may suppose that the fort is going to be taken and used against the enemy, and that it is therefore of importance to provide shot of the right calibre. One sailor says the calibre was eight inches and the other reported it was nine.

Clearly the decision to produce shots with diameter $8\frac{1}{2}$ inches makes no sense at all, since this contradicts the observations of both.

A good choice for the loss function in this example is

$$L = \begin{cases} 0 & \text{if} \\ 1 & \text{otherwise} \end{cases}$$

For such a loss function the arithmetic mean is usually not admissible. In this respect the situation is similar to Edgeworth's paradox.

Example 2. A wine-producing cooperative has received from a wine-grower a delivery of 1000 kilograms of each of the following variety of grapes: Veltliner, Traminer, Rotgipfler and Zweigelt. The sugar content of the grapes is determined by picking up a bunch of grapes from each variety. The rates of sugar content are the following: 1,5; 3,0; 2,5; 1,0. The vector as a result of this is inadmissible. An admissible estimator brings to the rates 1,33; 2,67; 2,38; 0,89.

Explanation. This phenomenon is called Stein's Paradox and it is based on the following assumptions. A sample of n independent normally distributed random variables $X_1, \dots, X_n$ with unknown means $\theta_1, \dots, \theta_n$ and known variance σ^2 is observed. The loss function of the estimator $T=(T_1, \dots, T_n)$ for $\underline{\theta}=(\theta_1, \dots, \theta_n)$ is quadratic, i.e.

$$L(T, \underline{\theta}) = \sum_{i=1}^n (T_i - \theta_i)^2.$$

For $n=1$ the estimator $T=X_1$ is optimal. For $n=2$ $T=(X_1, X_2)$ is at least admissible. If $n \geq 3$ the estimator $T=(X_1, \dots, X_n)$ fails to be admissible. There is a uniformly better estimator, namely

$$T = \left((1 - \sigma^2 \frac{n-2}{\sum X_i^2}) X_1, \dots, (1 - \sigma^2 \frac{n-2}{\sum X_i^2}) X_n \right)$$

(James - Stein estimator). This improved estimator is formed by multiplying the naive estimation vector $(X_1, \dots, X_n)$ with a constant smaller than one and thus by reducing the norm of the naive estimator. But this is most paradox: Since the origin is completely arbitrary, it can be replaced by any other point provided that this point is chosen in advance. By introducing some arbitrariness, the estimator is improved!

For an expert in estimation theory the phenomenon can be explained by the fact, that all admissible estimates are Bayes-estimates and that the James-Stein estimator is a Bayes-estimator for the normal distribution with mean zero as prior. This is however rather little intuitive. Therefore the opinion, that a similar phenomenon can only happen for dimension $n \geq 3$ and the quadratic loss function prevailed for a long time. This is one more argument for the above cited assertion that our intuition is trained by simple one- or two-dimensional examples and that we intend to transfer their simple features to more general situations. BERGER has however, given a full variety of cases (cf. BROWN 1980) in which non-admissibility holds even if $n \geq 2$. From the today's knowledge the admissibility of the intuitive estimator $(X_1, \dots, X_n)$ seems to be rather the exception than the rule.

Example 3. The location parameter (mean value) θ of a bimodal symmetric distribution is to be estimated. It is known that the two modes of the density are at distance 2 one from the other. A sample (X_1, X_2) of size two is observed. The efficiency of two different estimators is to be compared. We take

$$T_1 = \frac{X_1 + X_2}{2} \quad \text{and} \quad T_2 = \min(X_1, X_2) + 1$$

It may happen that

$$P\{(T_1 - \theta)^2 < (T_2 - \theta)^2\} < 1/2$$

(i.e. T_2 lies with higher probability closer to the true value) and at the same time

$$E((T_1 - \theta)^2) < E((T_2 - \theta)^2).$$

(i.e. T_1 has smaller risk). Which estimator is preferable?

Explanation. In order to simplify the explanation it is assumed that the random

variables X_i ; $i = 1, 2$ are discrete with distribution

$$X_i = \begin{cases} -1 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}$$

Then $P\{T_1^2 < T_2^2\} = 1/4$ but $E(T_1^2) = 1/2 < 1 = E(T_2^2)$.

Analogous examples with random variables possessing a density can be constructed in a similar manner. In particular the probability that the estimator with lower risk deviates more from the true value can be arbitrarily close to one. The reason for this paradoxon is simply the fact that the marginal distributions of a two dimensional random variables do not at all determine the common distribution. A further example in this direction is helpful:

It may happen that $P\{L_1 > L_2\}$ is (arbitrarily) close to 1 although L_1 is stochastically smaller than L_2 (i.e. $P\{L_1 \leq a\} > P\{L_2 \leq a\}$ for all a).

BLYTH (1972) gives the following example: The common density function $f(x, y)$ of L_1 and L_2 within the unit square is given by

$$f(x, y) = \begin{cases} 1/\epsilon & 0 \leq y - x \leq \epsilon \\ 1/\epsilon & x - y > 1 - \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Then $P\{L_1 > L_2\} = 1 - \epsilon$ but L_1 is stochastically smaller than L_2 . BLYTH describes also an interesting application: By observing two runners, who always run together, only one separate from the other, nothing can be said about the presumable winner of the contest. It may happen that runner 1 has a lower mean running time although runner 2 is with very high probability the winner in direct comparison. Transferring this example into the framework of WALD's statistical decision theory we can say that risk minimal estimators could be inferior in pairwise comparison. Thus some care is needed for the formulation of optimality criteria.

Similar paradoxa come from the non-transitivity of the relation $P\{L_1 < L_2\} > 1/2$. It may happen that the relations $P\{L_1 < L_2\} > 1/2$, $P\{L_2 < L_3\} > 1/2$ and $P\{L_3 < L_1\} > 1/2$ hold at the same time. Such examples are well known in the context of voting-paradoxa. In knockout tournaments

such paradoxical situations can arise as well (cf. ISRAEL (1981)).

The choice of the best among three alternative estimators T_1 , T_2 and T_3 seems to be ambiguous, if the corresponding loss functions L_1 , L_2 and L_3 exhibit such an intransitive behavior. Evidently, risk minimality is better than a criterion based on pairwise comparisons since the latter can lead to inconsistencies.

Example 4. A window was smashed when a burglary has been committed in a jewellery. A piece of broken glass was found in a suspect. The breaking index of the shop's window-pane was estimated to 1,518458. The breaking index of the glass splinter found in the suspect was investigated and calculated to the rate of 1,518472 with an standard deviation of 0,000004 due to the measurement error. A classical statistical test rejects the hypothesis of equality. Since the breaking indexes of window-panes vary between 1,51 and 1,53 the hypotheses of equality can be rejected from a bayesian standpoint, and thus the culprit could be caught. Which conclusion is the correct one?

Explanation. This phenomenon is called LINDLEY's Paradoxon. It can be described as follows: It may occur that a hypothesis is rejected with high significance by a classical test and is doubtless accepted by a bayesian standpoint. If, for instance, X is normally distributed with unknown mean θ and known variance σ then the hypothesis $\theta = \theta_0$ is rejected if $|X - \theta_0| > 4\sigma$ (say). If however a normal distribution with variance τ is taken as prior for θ and $\tau \gg \sigma$, than X may lie in the acceptance region. The paradox conclusion is: The more vague the prior information is, the more diffuse the prior distribution is to be chosen and the more likely is the acception of the null hypothesis. Or to put it into the language of the example: The less there is known about the distribution of the breaking indices of glass, the more likely is the guilty of the suspect.

The different positions (classical versus bayesian standpoint) can be characterized as follows: The attorney uses the argument: The breaking indices of the two glasses are such different one from the other, compared to the error of measurement, that it is impossible that the suspect had committed the crime. The reply of the public prosecutor is: There are so different types of glass windows in our town that it is rather unlikely that the suspect had a piece of

glass in his pocket, which is so similar to that in the jewellery. How to solve this dilemma?

SHAHER (1982) utilizes the theory of belief-functions, which was invented by himself, to resolve the problem. His conclusion is that the prior information on the distribution is less relevant than the measurement. LINDLEY comments on that by saying that SHAHER was led by a subtle form of base-rate fallacy (TVERSKY & KAHNEMANN 1974). From LINDLEY's viewpoint the bayesian standpoint is superior to the non-bayesian if the prior information is empirically based.

There is no general way out of the dilemma. To the contrary: The question of the use of prior probabilities in statistical estimation separates the scientists in two disjoint groups. The decision, to which group one adheres, eliminates the paradoxon on the individual level. To the author's opinion however, this way out of the contradiction is not quite satisfactory.

References

- Blyth, C.R.: 1972. Some probability paradoxes in choice from among random alternatives. In: J. Amer. Statist. Assoc. 67, 366-373
- Brown, L.: 1980. Example of Berger's phenomenon in the estimation of independent normal means. In: Ann. Statist. 8, 572-585
- Edgeworth, Y.: 1983. The law of error. In: Philosophical Magazine 16, 300-309
- Israel, R.B.: 1981. Stronger players need not to win more knockout tournaments. In: J. Amer. Statist. Assoc. 76, 950-951
- James, W. & Stein, Ch.: 1961. Estimation with quadratic loss. Proceedings of the 4th Berkeley Symposium, Vol. 1, 361-379
- Lindley, D.V.: 1977. A problem in forensic science. In: Biometrika 64, 207-213
- Pflug, G.Ch.: 1980. Paradoxien der Wahrscheinlichkeitsrechnung. In: Stochastik im Schulunterricht. Hölder, Pichler, Tempsky, Wien
- Scholz, R.W.: 1981. Stochastische Problemaufgaben - Analysen aus didaktischer und psychologischer Sicht. IDM Materialien und Studien Bd. 23, Bielefeld
- Shafer, G.: 1982. Lindley's paradox. In: J. Amer. Statist. Assoc. 77, 325-334
- Stein, C.: 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the 3rd Berkeley Symposium, Vol. 1, 197-206

- Stigler, S.: 1980. An Edgeworth curiosum. In: *Ann. Statist.* 8, 931-934
- Tversky, A. & Kahneman, D.: 1974. Judgement under uncertainty: Heuristics and biases. In: *Utility, Probability, and Human Decision Making*, 141-162
- Wald, A.: 1950. *Statistical decision functions*. New York: John Wiley

This Page Intentionally Left Blank

COMMENTS

This Page Intentionally Left Blank

THE POSSIBLE AND THE PROFITABLE

Some remarks on the symposium's impact on decision research

BARUCH FISCHHOFF

Decision Research
A Branch of Perceptronics
USA

One common response to feelings of unease about the current state in a field research is to search for a best method for the future. In this meeting, we have heard a variety of persuasive and inconsistent arguments for how the study of individual decision making should proceed. Projecting from my own response, these appeals have fallen on open minds, but un-receptive hearts. The ensuing discussions have shown great willingness to take other people's results seriously, but little readiness to shift one's own research program. I suspect that such changes of course are as much a function of constitution as of conviction. Some of us like to ask people a lot of questions face to face; others like to keep subjects at arms' length. Some of us like to proceed by little steps; others like to plot grand re-search strategies. Some prefer the risk of creating theories that cannot be falsified; others prefer the risk of producing empirical results of unknown generality.

If we are to have anarchy, rather than just chaos, then we need to know how best to learn from one another and to exploit one another's loosely coupled experience. One distinction that helps me to understand the relevance of the diverse work described here to my own work is between top-down and bottom-up approaches. The top-downers like to lay out a complete frame of reference including terms, methods, theories, assumptions, and domain of in-

quiry, before collecting any data (if there is any time left to do so or any experiment that will meet all these criteria). The bottom-uppers like to discover solid, replicable effects that can be accounted for by modest local theories, before worrying about formal definitions for the semi-intuitive concepts that they use (if there is time left to do so and a clear correspondence can be made between their experiment and any theory).

One possible analogy is with the differences between those who try to understand a forest by aerial photography and those who try to do so by exploring clearings on the ground. Like the botanists who have used these contrasting procedures, we tend to find that forests with a full canopy often do not look so healthy on the ground. There may be little regeneration, parasitic diseases in incipient stages, or soil compactment that are hidden by the foliage of a standing forest. Conversely, the problems seen in one locality may not be found widely, may not threaten the forest as a whole, and may even be necessary for the stability of the entire ecosystem.

For those of us working on the ground, the critical question is where the trees that we know so well fit into the overall forest. It is always easy to speculate about the boundary conditions for any observed phenomenon. Classic ad hoc claims are that the phenomenon would disappear if the question were posed differently, if subjects were better motivated, or if subjects could be induced to adopt the experiment's view on how the problem should be structured. A bottom-up strategy for addressing such claims is to review available studies to see whether any of these additional factors consistently makes a difference (which would give it the status of an explanatory rather than a nuisance variable).

A more top-down procedure is to compare notes with those who study different strata on the forest or aspects of its functioning. A much stronger claim can be made for the necessity of reconciling one's own results with other robust results than for taking seriously mere speculations about what else might be happening. In this regard, the conference has been extremely productive. From the point of view of research into individual decision making behavior, we have seen a variety of useful alternative cuts at our

common problem. One is the study of the ways and frequency with which group situations can interfere with individual rationality. Another is the difficulties that arise when the demands of a decision-making fail to match the decision maker's abilities or organizational setting. A third is some of variety of alternative ways in which subjects might reasonably (mis)-construe our stimuli. A fourth is the kind of educational and developmental experiences that may have shaped the cognitive skills that subjects bring to our tasks. A fifth is the kind of judgmental difficulties that have been found even with people who should be experts and with problems that should demand the best performance of which they are capable. Our task now is to go back home and quietly digest this material, in order to see which of these possibilities are actually probabilities.

This Page Intentionally Left Blank

THE IMPORTANCE OF THE FICTIONAL FOR THEORIES OF KNOWLEDGE

Some remarks on the symposium's impact on cognitive psychology

ALF C. ZIMMER

*Universität Münster
Psychologisches Institut
F.R. Germany*

The quest of Tversky, Kahneman, and others for persistent biases in human judgment (for overviews see KAHNEMAN, SLOVIC, and TVERSKY, 1982, or NISBETT and ROSS, 1980) is reminiscent of HELMHOLTZ' (1896) approach to the investigation of human perception: in both cases discrepancies between the "real" on the one hand and the "perceived" or "judged" on the other hand have been used as starting points for the investigation of the differences between human information processing and the way the world is described in terms of physics or logic. HELMHOLTZ' analysis of optical illusions influenced him to postulate the theoretical construct "unconscious inferences" which in turn led to the interpretation of perception as the result of processes analogous to computation (PYLYSHYN, 1980). The approach to human decision making taken by TVERSKY and others follows an analogous research program by investigating under which circumstances human (unaided) judgment fails either to process all the necessary aspects of information available or to represent them veridically.

GIBSON (e.g. 1979) has challenged the HELMHOLTZian paradigm by pointing out that from an evolutionary point of view it is not the mismatches between the given and the perceived that are of interest but the matches. That is, the veridical usage of the affordances provided to us by our environment are of interest. I have suggested (ZIMMER, note 1) that these seemingly opposing

views can be regarded as different aspects of the same process, in one case from a bottom-up perspective and in the other case from a top-down perspective. Both aspects seem to be necessary for a complete understanding of complex perceptual or judgmental tasks. Complementary to the approach taken by TVERSKY and others, I propose to take the conditions for veridical judgments as a framework in which biases can be traced back to the conditions under which such judgments are ecologically valid. (see LOPES, 1982)

In order to develop a theory of knowledge in this framework one has to account for an aspect of human knowledge that is usually disregarded: the capacity to generate virtual experiences by using idealized models of the reality, which are more or less fictional. This fictional aspect of knowledge can be used in order to sketch future outcomes of decisions or to derive other consequences. The matching of such virtual experiences derived from fictional knowledge in turn seems to underly what has been termed the expectancy effect in perception. That is, what we see is strongly influenced by expectations we have about the regularities of our environment, which are at least partially fictions of our mind. Relying on these expectancies allows for a very fast processing of information. The cost for processing information so fast is that there sometimes are deleterious side effects, which become especially apparent in the tricks magicians play (GREGORY, 1970, MULHOLLAND, 1927), where the audience is tricked into expecting an action, which does not happen, but which works as a camouflage for what is really happening. Virtual experiences due to the fictional aspects of knowledge may influence the selection of conjectures and predictions. Virtual experiences may also play an important role in hindsight explanations.

These ideas follow in some respect VAIHINGER (1911) who in his "Philosophy of "as if" developed a theory of knowledge founded on "theoretical, practical, and religious fictions of mankind". He showed how knowledge can be expanded by pretending "as if". For instance "Let's pretend as if there were two truly parallel lines" or "Let's pretend all the mass of a body were in its center of gravity." As we know these "as if" pretensions have been crucial for the development of geometry or physics, were idealizations abound. What is missing in VAIHINGER's theory is a connection between fictional knowledge and knowledge due to experiences. Theories on the subjective assessment of one's

own knowledge (How good are my reasons for assuming that a given idealization or a fictional scenario turns out to be an apt tool for describing or predicting events?) might be able to bridge this gap.

Such subjective assessments of the fictional aspects of knowledge can be done in terms of possibility (ZADEH, 1978), plausibility (RESCHER, 1973), or subjective probability. Therefore they should be accounted for by models for human information processing and theories of action. I have argued (ZIMMER¹) that it is plausible to assume that this "knowledge about knowledge" is represented internally by means of a propositional language-like code. The interpretation of verbal expressions for this knowledge plays therefore an important role in the understanding of such elements of conversation which cannot be reduced to demonstrative procedures and consecutive observations.

Up to this point only the individual knowledge has been taken into account but much of the knowledge used in mastering the environment is not due to the individuals' own experiences but is due to vicarious experiences made by other people and shared afterwards. How the necessity to share knowledge imposes constraints upon information processing has been investigated in shareability theory (FREYD, in press). One important consequence of this theory for the investigation of the effectiveness or biasedness of human judgment is that it forces us to distinguish between the effects due to a lack of knowledge (content and procedures) and those due to the fact that certain real and virtual experiences do not follow shareability constraints. The kind of repairs in conversation used in order to overcome the misunderstandings caused by these effects might provide an empirical approach to this problem. For instance if one assumes that a misunderstanding is due to a lack of knowledge then an appropriate repair would be to give a definition, on the other hand if the misunderstanding is assumed to be due to disregarding shareability constraints an appropriate repair could consist in suggest-

¹see ZIMMER, A. Verbal vs. numerical processing of subjective probabilities, in this volume.

ing: "Imagine a situation where the following is true: ...".²

Reference Notes

1. What makes the eye intelligent? manuscript, Stanford University, Department of Psychology, 1981

References

- Freyd, J.J.: Shareability: the social psychology of epistemology. In: Cognitive Science, in press
- Gibson, J.J.: 1979. The ecological approach to visual perception. Boston: Houghton and Mifflin
- Gregory, R.L.: 1970. The intelligent eye. New York: MacGraw-Hill
- Helmholtz, H. von: 1896. Handbuch der physiologischen Optik (2nd ed.). Hamburg und Leipzig: Voss
- Kahneman, D./Slovic, P. & Tversky, A. (eds.): 1982. Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press
- Lopes, L.L.: 1982. Doing the impossible: A note on induction and the experience of randomness. In: Journal of Experimental Psychology / Learning, Memory, and Cognition 8, 626-636
- Mulholland, J.: 1927. Quicker than the eye. Indianapolis: Bobbs-Merrill
- Nisbett, R. & Ross, L.: 1980. Human Inference: Strategies and Shortcomings of Social Judgments. Englewood Cliffs: Prentice Hall
- Pylyshyn, Z.: 1980. Computation and cognition: Issues in the foundation of cognitive science. In: The Brain and Behavioral Science 9, 111-169

² These rather theoretical considerations about the connection between knowledge and subjective probabilities permit the deduction of educational consequences concerning the use of fallacies in human judgment as a means of teaching statistics or decision theory. Among others one beneficiary effect might be the demonstration that immediate impressions tend to be misleading because of the expectancy effect. But inducing students to produce judgmental fallacies might have unwanted negative effects too, because students might experience this behavior as a kind of non-cooperative discourse. This in turn might lead to the consequence that it is not the sensitivity for matches vs. mismatches of one's own virtual experiences with real situations that is heightened. It might be that instead of this merely the vicarious experience is made how somebody can be tricked into inconsistency.

Rescher, N.: 1973. *The Coherence Theory of Truth*. Oxford: Clarendon Press

Vaihinger, H.: 1911. *Die Philosophie des Als-Ob*. Berlin: Reuther und Reichard

Zadeh, L.: 1978. Fuzzy sets as a basis for a theory of possibility. In: *Fuzzy Sets and Systems* 1, 3-28

This Page Intentionally Left Blank

WHY ARE DEVELOPMENTAL PSYCHOLOGISTS BIASED WITH REGARD TO THE BIASEDNESS OF THE HUMAN SUBJECT?

Some remarks on the symposium's impact
on developmental psychology

MANFRED WALLER

*Universität Bielefeld
Abteilung für Experimentelle und Angewandte Psychologie
F.R. Germany*

That human judgements are biased, as impressively substantiated by many of the contributions to this conference, is a truism for the developmental psychologist. He is almost daily confronted with it in manifold forms when he investigates the growth of knowledge and judgement structures in the growing child, for example, in the form of infantile egocentrism (PIAGET, 1955).

That this also applies in similar ways for the judgements made by the adult, fully developed, subject (for a review c.f. KAHNEMAN, SLOVIC & TVERSKY, 1982) may, on the other hand, be a somewhat surprising result of this conference for the developmental psychologist. It presents him with a challenge, for it seems to contradict one of his implicit basic assumptions, namely the assumption that, in the course of his development, the growing child surmounts step by step the obvious deficiencies and biases of his judgement ability, and at the completion of his development becomes an almost perfect being, a logically and rationally (i.e. unbiased) thinking *quasi-scientist* as PIAGET (1970) postulates in his cognitive developmental theory. This similarly applies for the development of other functions; for example, speech development according to CHOMSKY (1965) and moral development according to KOHLBERG (1976).

Regarding the evident contradiction between this view and the main findings of the conference, the developmental psychologist has to consider SHAKLEE's (1979) critical question, whether the postulate of a quasi-unlimited developmental progression can be maintained.

In order to answer this question adequately, it is first necessary to point out that there is a basic difference between the biased judgement of a child, such as the non-recognition of the invariance of a quantity of liquid that is poured from one container into another, and the false judgement of an adult based on judgement distortions or judgemental heuristics (c.f. KAHNEMAN & TVERSKY, 1982). A child's judgement bias is generally a question of competence. This means that before a definite stage of development is reached, a child is basically not able to solve an invariance problem because of specific limitations of his cognitive system (for example, lack of knowledge, limitations of working memory and of attentional resources, etc.). Additionally, because of the non-reflexivity of his thinking, he is not capable of becoming aware of the one-sidedness and bias of his false judgement.

However, the biased judgement of the adult subject has to be assessed differently. It is generally not due to a subject's lack of competence, but is a question of performance conditions. In other words, the adult is basically able to correct his error given suitable task conditions, especially because he has the ability to recognize the incorrectness of his judgement and critically reflect upon it.

Because of this, it seems to be problematic, and for a developmental psychologist unjustifiable, to attribute children's judgement bias to the same judgement heuristics as those of adults, as ROSS (1981) for example, attempts. ROSS refers to findings reported by JOSEPHSON (1977, quoted by ROSS, 1981) who was able to show that in the prediction of situation-specific behaviour of other persons, 5 year old children place greater emphasis on situative factors than on dispositional ones, just as 20 year olds do, and that they therefore make better behaviour predictions than a group of children of a mean in-between age who put more emphasis on dispositional factors. According to ROSS these results indicate that the judgements of the 5 year olds are (as yet) less distorted by the "fundamental attributional error" than those formed by older children. Such an interpretation is inadmissible. It doesn't take into account that pre-school age children exclusively take notice of external, i.e. situative and/or physical or accessory features, and that they are still generally blind to dispositional features such as traits and the like. This is clearly shown by research into the development of person perception (for a review, c.f. LIVESLEY & BROMLEY, 1973, and SECORD & PEEVERS, 1974).

It would therefore be wrong to draw the conclusion from the biased judgments of adult subjects that they have made no progress in their cognitive development when compared with the erroneously judging child. Developmental psychological research has repeatedly and unanimously shown what these developmental advances are. They consist on the one hand in the enlargement of specific competencies and skills (e.g. in the enlargement of the effective capacity of working memory, in the acquisition of knowledge and control strategies, etc.), and on the other hand, in the growing ability to control and critically reflect upon his own thinking as a result of his meta-cognitive development (c.f. BROWN & DeLOACHE, 1978). Certainly, however, such developmental advances obviously do not rule out the judgement errors made by adults.

This is very clearly documented by the results of this conference, but is also well known to the developmental psychologist. In particular, cross-cultural studies (in summary c.f. COLE & SCRIBNER, 1974) reveal that it is only under optimal development and socialization conditions that the adolescent subject attains the state of development that CHOMSKY (1965), KOHLBERG (1976), PIAGET (1970) and others, consider to be the final stage of the developmental progression. In addition to this, interindividual differences must also be taken into account.

However, in their research practice, developmental psychologists often pursue a strategy that aims at fully exploring the capabilities and competencies of the growing subject by investigating the abilities of children or young people under optimal if not *ideal* performance conditions (c.f. SHAKLEE, 1979). In developmental psychological research, the subjects being investigated are generally prepared for their task with great care, and are therefore, as a rule, highly motivated. The problems they have to solve can often be worked on without time limits and also, due to the results of more or less detailed task analyses, these problems are mostly presented in such a form that all informations relevant for their solution are available to the subject.

Such performance conditions, that so favour a subject's responses, are seldom given for the solution of everyday problem situations, at least, not as a rule. Additionally, the scientific interest of developmental psychologists is too often, as for example in the case of PIAGET, principally aimed at general or modal development trends, with the consequence that interindividual differ-

ences are lost from view.

Both the specific scientific interest of the developmental psychologist, and the particular features of his research strategy, may result in the growing subject becoming an *ideal* subject within the realms of developmental psychology: a subject who due to the neglect of interindividual differences, and by the implicit assumption of optimal developmental conditions, tends to become a *quasi-scientist*, with unlimited and unbiased judgement capability. Thereby the ecological validity of the results of developmental psychology research may be greatly reduced.

This conference can draw the attention of developmental psychologists to exactly this problem, and hopefully stimulate them to investigate the developing capabilities of the growing subject under conditions that place more emphasis on ecological factors.

References

- Brown, A.L. & DeLoache, J.S.: 1978. Skills, plans, and self-regulation. In: R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum, 3 - 35
- Chomsky, N.: 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press
- Cole, M./ Scribner, S.: 1974. *Culture and thought: A psychological introduction*. London: Wiley
- Josephson, J.: 1977. The child's use of situational and personal information in predicting the behavior of another. Unpublished doctoral dissertation. Stanford: Stanford University
- Kahneman, D. , Slovic, P. & Tversky, A. (Eds.) : 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press
- Kahneman, D. & Tversky, A.: 1982. Judgment under uncertainty. In: D. Kahneman/ P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press, 3 - 20
- Kohlberg, L.: 1976. Moral stages and moralization. The cognitive-developmental approach. In: Th. Lickona (Ed.), *Moral development and behavior. Theory, research, and social issues*. New York: Holt, Rinehart and Winston, 31 - 53

- Livesley, W.J. & Bromley, D.B.: 1973. Person perception in childhood and adolescence. London: Wiley
- Piaget, J.: 1955. The language and thought of the child. Cleveland: Meridian Books
- Piaget, J.: 1970. Piaget's theory. In: P.H. Mussen (Ed.) Carmichael's manual of child psychology. Vol. 1. New York: Wiley, 703 - 732
- Ross, L.: 1981. The "intuitive scientist" formulation and its developmental implications. In: J.H. Flavell & L. Ross (Eds.), Social cognitive development. Frontiers and possible futures. Cambridge: Cambridge University Press, 1 - 42
- Secord, P.F. & Peevers, B.H.: 1974. The development and attribution of person concepts. In: T. Mischel (Ed.), Understanding other persons. Oxford: Blackwell, 117 - 142
- Shaklee, H.: 1979. Bounded rationality and cognitive development: Upper limits on growth? In: Cognitive Psychology 11, 327 - 345

This Page Intentionally Left Blank

INDIVIDUAL DECISION MAKING AND SOCIAL PSYCHOLOGY

Some remarks on the symposium's impact on social psychology

KATRIN BORCHERDING

*Universität Mannheim
Sonderforschungsbereich 24
F. R. Germany*

The area of social psychology can roughly be defined as 'investigating and explaining individual behavior in social contexts'. Since 'individual decision making under uncertainty' is increasingly analysed in social situations, it becomes also a field of social psychology. Close interrelation between parts of social psychology and decision research shall be demonstrated by two topics which were discussed during the symposium, namely distributional decisions using the classical paradigm of n -person games (see the contributions by ALBERS, by SCHULZ & MAY and by TACK, this volume) and consensus decisions with the classical finding of choice shift (see CROTT & ZUBER, this volume). Both topics have been treated by many social psychologists as well as decision researchers; they are therefore suited to point out parallels as well as typical differences in perspectives.

Distributional decisions

Distributional decisions are central to individuals, groups and societies as a whole in that available, usually limited resources or 'goods' like time, money, rights, etc. have to be allocated or shared. The substitutability of goods is investigated in decision research, e.g. in the area of multi-attribute utility theory (KEENEY & RAIFFA, 1976), and in social psychology, e.g. in theories of social exchange using economic principles to explain social interaction (HOMANS, 1961).

In decision research distributional decisions are studied with a focus on individual aspects of the situation: In game theory, where matrices specify the outcomes for the various players, individual behavior is analysed as a function of matrix outcomes as being e.g. rational, competitive or cooperative. In bargaining decisions, where money is shared by the players according to some rules, individual behavior is analysed in order to learn out peoples' strategies to achieve their goals and use their power. The social impact of the situation is not stressed and the necessary interactions between the players are reduced to a minimum and often done using technical equipments without personal contact.

In the area of social psychology distributional decisions are analysed according to whether they are evaluated as just and how people react to violations of distributive justice. When talking about an adequate partition of a common good, three principles of justice are discussed: partition according to contributions (equity principle), partition according to needs (Marxian principle) and equal partition where everyone gets the same (parity principle). Some authors justify the equity principle as universal (e.g. WALSTER, WALSTER & BERSCHIED, 1978), others argue that different principles are appropriate in different situations (DEUTSCH, 1975; KAYSER & SCHWINGER, 1982; MIKULA, 1980; SCHWINGER, 1980), and LEVENTHAL assumes that all three principles enter into a fair solution, perhaps weighted differently (LEVENTHAL, 1976). If several principles apply either simultaneously or alternatively the question arises for the conditions under which a specific principle is used or - in the LEVENTHAL model - for the factors that influence the relative weights. The extent to which a principle is evaluated as just might depend on (a) the kind of task for which the common 'payment' is done (intellectual or not, easy dividable or not, time consuming or not; a relevant taxonomy of task characteristics has yet to be developed); (b) the kind of resource in which the 'payment' occurs (money, time, love, status, etc.; see FOA & FOA, 1976, for a theory of resources); (c) the relations between the subjects (friends, relatives, neighbours, strangers, etc.; see HASSEBRAUCK, 1982, for some results); and (d) the social context in which all this takes place (work, leisure, etc.). Thus, the same question - how do people partition a common good - is approached in a formal and structural way in decision research and in a more substantive way by social psychologists.

Consensus decisions

The first publications on choice shift (STONER, 1961, 1968; KOGAN & WALLACH, 1964) led many social psychologists to investigate the question whether group judgments differ from individual judgments. They found systematic differences, which they called 'shifts', and depending on the direction there are 'conservative' and 'risky' shifts. Since many decisions are done in groups or committees, effects of group behavior on decisions are relevant and knowledge about causes and consequences of such shifts is needed. The procedure in choice shift research is (a) elicit individual judgments, (b) aggregate them, (c) observe group judgments, (d) compare group judgments with the aggregation of individual judgments, and (e) explain differences. Since group judgments are interpreted in relation to an aggregation of individual judgments, the adequacy of the aggregation algorithm becomes crucial; the relevance of the findings depend on the extent to which this aggregation can serve as a 'rational' procedure. In almost all cases - without giving reasons and hardly ever questioned- some arithmetic mean was taken as the aggregation algorithm (but see as exceptions BORCHERDING & KISTNER, 1982; CROTT & ZUBER, this volume; JEAN, 1970; and SCHAEFER, 1978), and many theories are developed to explain the results (for an overview see SAUER, 1974; SIX, 1981). As CROTT & ZUBER have shown, there are alternative aggregation procedures which are at least as reasonable as others but wash out the existence of shifts. Social psychologists working in the area of choice shift or polarisation should therefore concentrate more on the appropriateness of aggregation procedures and if shifts still occur, there is a need for theories which can predict the direction of a shift as a function of task characteristics for a variety of tasks. It is remarkable that consensus behavior was never analysed on an individual level. Knowledge about how easy statements can be accepted by an individual might be fundamental for consensus in groups. Determinants of the 'individual acceptability of divergent statements' might be - among others - the importance of the decision situation for an individual, the direction and/or magnitude of the discrepancy from own opinions, and the reversibility of the potential consequences. Such research could be crucial to determine whether choice shift is a phenomenon that occurs only in groups or whether it is already established firmly in individual behavior.

Outlook

General conclusions are difficult and only preliminary because of the breadth of the areas. Sometimes it appears as if the strength and weakness of a discipline lies closely together. Social psychologists have investigated individual behavior in a variety of situations and looked for the underlying determinants of the respective behavior. The development of substantive theories was promoted in order to efficiently and economically explain the variety of observations and results. But now and again social psychologists may be somewhat hasty to jump at conclusions and to invent new theories without sufficiently considering the fundamentals in measurement, the specificity of the situation, alternative causes, and conflicting theoretical explanations. Social psychology is an area with relevant problems that need solutions; but social psychologists are sometimes more committed to confirm their theories than to solve the problems.

In comparison decisions research is more oriented towards the solution of real problems, but decision researchers sometimes narrowly focus on the formal structure of a situation and on measurement procedures without sufficiently considering the meaningfulness of their variables, the psychological co-determinants and substantive theoretical explanations of the decision behavior. Since individuals' representation of the task is not perfectly overlapping with the one intended by the researcher, individual behavior is besides individual aspects always influenced by social ones as well. Evidence for the impact of social components on individual decisions is given by TACK (this volume). His persuasive 'egoistic' concept of 'individual rationality' was violated by his subjects. Analysing the situations where most of the violations occurred led him to conclude: "Individuals with more power not justified by achievement might violate individual rationality in taking social values into account besides their individual interests". Deviations from this 'individual rationality' in order to reduce existing interindividual differences is a behavior in accordance to perceived distributive justice. A higher order of individual rationality could incorporate social motives like the ones given by MACCRIMMON & MESSICK (1976).

Research in areas, where scientists from different fields with various backgrounds work on can be improved by a better communication between these re-

searchers. This is an old and somewhat trivial statement, but not so often realized as desirable. Symposia like this one are important and promising as issues get enlarged, and broader perspectives in research will improve the potential impact and applicability of its results.

References

- Borcherding, K. & Kistner, K.: 1982. Entwicklung und Überprüfung eines Modells zur Gruppeneinigung in Choice-Dilemma-Situationen. *Zeitschrift für Sozialpsychologie*, 13, 323-332.
- Deutsch, M.: 1975. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 31, 137-149.
- Foa, E.B. & Foa, U.G.: 1976. Resource theory of social change. In: J.W. Thibaut, J.T. Spence & R.C. Carson (eds.), *Contemporary topics in social psychology*. Morristown, N.Y.: General learning press.
- Hassebrauck, M.: 1982. Reaktionen auf distributive Ungerechtigkeit und Art der Sozialbeziehung. Eine Überprüfung des Mehr-Prinzipien-Ansatzes. Beitrag auf der 24. Tagung experimentell arbeitender Psychologen in Trier.
- Homans, G.C.: 1961. *Social behavior: Its elementary forms*. New York, N.Y.: Harcourt, Brace & World.
- Jean, R.St.: 1970. Reformulation of the value hypothesis in group risk taking. *Proceedings of the 78th Annual Convention, APA*.
- Kayser, E. & Schwinger, T.: 1982. A theoretical analysis of the relationship among individual justice concept, layman's psychology, and distribution decisions. *Journal of the Theory of Social Behavior*, 12, 47-51.
- Keeney, R.L. & Raiffa, H.: 1976. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York, N.Y.: Wiley.
- Kogan, N. & Wallach, M.A.: 1964. *Risk taking: A study in cognition and personality*. New York, N.Y.: Holt, Rinehart & Winston.
- Leventhal, G.S.: 1976. Fairness in social relationships. In: J.W. Thibaut, J.T. Spence & R.C. Carson (eds.), *Contemporary topics in social psychology*. Morristown, N.Y.: General learning press.
- MacCrimmon, K.R. & Messick, D.M.: 1976. A framework for social motives. *Behavioral Sciences*, 21, 86-100.
- Mikula, G.: 1980. Zur Rolle der Gerechtigkeit in Aufteilungsentscheidungen. In: G. Mikula (Hrsg.), *Gerechtigkeit und soziale Interaktion*. Bern: Huber.
- Sauer, C.: 1974. Zur Erforschung der Gruppenextremisierung nach Diskussion. *Zeitschrift für Sozialpsychologie*, 5, 255-273.

- Schaefer, R.E.: 1978. Eine entscheidungstheoretische Analyse des Risky Shift Phänomens. *Zeitschrift für Sozialpsychologie*, 9, 186-205.
- Schwinger, T.: 1980. Gerechte Güterverteilungen: Entscheidungen zwischen drei Prinzipien. In: G. Mikula (Hrsg.), *Gerechtigkeit und soziale Interaktion*, Bern: Huber.
- Six, U.: 1981. Sind Gruppen radikaler als Einzelpersonen? Ein Beitrag zum Risikoschub-Phänomen. Darmstadt: Steinkopff.
- Stoner, J.A.F.: 1961. A comparison of individual and group decisions involving risk. Unpublished master thesis. Cambridge, Mass.
- Stoner, J.A.F.: 1968. Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology*, 4, 442-459.
- Walster, E., Walster, G.W. & Berscheid, E.: 1978. *Equity theory and research*. Boston, Mass.: Allyn & Bacon.

TOWARDS A THEORY OF LIMITED RATIONALITY

Some remarks on the symposium's impact on economic theory

REINHARD SELTEN

*Universität Bielefeld
Institut für Mathematische Wirtschaftsforschung
F.R. Germany*

The results on decision making under uncertainty reported in this conference belong to a stream of accumulating evidence whose impact on economic theory may eventually turn out to be of great significance. The prevailing picture of economic man needs to be reformed in the direction of limited rationality. Unfortunately experimental research did not yet produce a sufficiently powerful theory of limited rationality. It is the purpose of this short paper to emphasize the necessity of further work towards this goal.

The prevailing picture of rational man in present day economic theory is that of a Bayesian decision maker with unlimited cognitive capabilities. Of course, economists admit that this is an idealization which does not completely correspond to the facts. Defenders of Bayesianism as a descriptive theory argue, that deviations from reality are relatively unimportant. Moreover, they maintain that even if individual behavior is far away from Bayesian rationality average behavior may still be correctly described, since deviations in opposite directions can be expected to cancel out. Utility maximization is defended for the "representative household", even if individual households are known to make intransitive choices. In my mind there is no doubt that such positions are not tenable in the face of the experimental evidence. Systematic biases and typical fallacies are observed and not just random deviations from Bayesian rationality.

In the last years more and more economists have become aware of the discrepancies between experimental facts and the conventional wisdom of economic textbooks. Phenomena like those reported in this conference cannot be overlooked any more. However, economic theory has been quite successful in spite of its unrealistic microeconomic foundations. Economists cannot afford to reject the traditional view unless it can be replaced by a better theory.

A great advantage of the prevailing neoclassical approach is the easiness with which it generates meaningful theoretical questions. If a new economic problem arises - say a sudden worldwide scarcity of food - readily available modelling tools can be used in order to describe the technological, institutional, and motivational features of the situation. Production functions, market structures, economic policy instruments and utility functions are combined to form a quantitative model which can be used to explore welfare implications of alternative economic policies.

The interests of those who have to make decisions are described by utility functions. Behavioral predictions are based on the principle of subjective utility maximization. Welfare measures are defined in terms of utilities. The role of the principle of subjectively expected utility maximization is a very important one. It is especially hard to see how questions concerning welfare implications of policy measures can be posed without any recourse to utility theory.

The experimental evidence on human decision behavior shows what is wrong with present day economic theory, but it does not yet show how to improve on it. It is necessary to develop a theory of limited rationality which can replace the principle of subjectively expected utility maximization as a tool of economic model building.

A useful theory of limited rationality will have to be nearer to psychological reality, but it may still be a very idealized picture of empirical phenomena. It should be simple enough to be incorporated into quantitative economic models amenable to analytical treatment. Obviously, one cannot avoid a compromise between realism and simplicity.

Up to now attempts to modify economic theory in the direction of limited rationality have not been as successful as one might wish. The computer simulation studies inspired by the behavioral theory of the firm (CYERT & MARCH 1963) and by the evolutionary approach (NELSON & WINTER 1974) have produced admirable results, but they failed to have a major influence on the main body of economic theory. Generally, economists feel uneasy about complicated computer simulations; often it is hard to see to which extent the results depend on specific parameter values and other ad hoc features of the simulation model. Nevertheless computer simulation studies may be an important source of interesting new economic effects which then can be studied by means of analytic models.

Following the seminal work of SIMON the author has participated in an attempt to develop a general theory of economic decision making under limited rationality (SIMON 1955, SAUERMAN & SELTEN 1962). However, this approach was highly speculative and only loosely based on casual empirical observations. The basic idea of aspiration level adaptation has proved to be fruitful in experimental research, but the theory in its original form cannot claim to be an explanation of empirical facts. Mere philosophical speculation is not a reliable basis for the development of a theory of limited rationality. It is necessary to gain insight on real human decision behavior. In this respect experimental research is the most important tool.

Already in the late fifties a small group of economists felt that they had to do experimental research in order to answer questions arising in the theory of the firm. In Germany this has led to a series of volumes "Beiträge zur experimentellen Wirtschaftsforschung - Contributions to experimental economics". The movement of experimental economics has come about, since in the time of its beginnings psychologists did not yet pay much attention to behavioral problems of economic theory. Nowadays this has changed and it is fair to say that most work in experimental economics is done by psychologists.

Much of the experimental work on human decision making strongly emphasizes biases and fallacies. In my view it is wrong to criticize this tendency as the expression of an unduly negative attitude towards human cognitive capabilities. Biases and fallacies probably are the key to a better understand-

ing of the marvelous information processing methods which enable human beings to find practical solutions to problems whose complexity defies any attempt to apply Bayesian decision theory.

Undoubtedly the further development of decision psychology is not only of great theoretical importance but also of considerable practical significance. Without insight into the structure of natural decision behavior we cannot hope to learn how complex decision tasks like computer administrated business games should be approached by a decision maker subject to the usual cognitive constraints imposed by time pressure and task complexity. Of course, in order to answer such questions we need a theory of limited rationality which supplies quantitative answers rather than qualitative ones. It is not sufficient to know whether a price should be raised or lowered a businessman has to decide by how much to change the price.

The results of the conference show that experimental research on human decision making is making great progress towards a better understanding of limited rationality. Everyone who works in the field should be aware of the importance of the subject for economic theory.

References

- Cyert, R.M. & March, I.G.: 1963. A behavioral theory of the firm. Englewood Cliffs, N.J.: Prentice Hall
- Nelson, R.R. & Winter, S.G.: 1974. Neoclassical vs. evolutionary theories of economic growth: critique and prospectus. In: *The Economic Journal* 84, 886-905
- Simon, H.A.: 1955. A behavioral model of rational choice. In: *Quarterly Journal of Economics* 69, 99-118
- Sauermann, H. (ed.): 1967, 1970, 1972, 1973, 1974, 1976, 1978. Beiträge zur experimentellen Wirtschaftsforschung - Contributions to Experimental Economics, Volumes I to IX. Tübingen: I.C.B. Mohr
- Sauermann, H. & Selten, R.: 1962. Anspruchsanpassungstheorie der Unternehmung. In: *Zeitschrift für die gesamte Staatswissenschaft* 118, 577-597

TOWARDS AN EDUCATION FOR DECISION BEHAVIOR

Some remarks on the symposium's impact on mathematical education

HANS-GEORG STEINER

*Universität Bielefeld
Institut für Didaktik der Mathematik
F.R. Germany*

Before I try to say something about a possible impact of the conference on mathematical education, I would like to give a brief view on mathematical education as I see it interpreted and operationalized by the *Institute for the Didactics of Mathematics* (IDM) which through Dr. Scholz signs as the originator of the conference. As you might have seen from the last pages of the book on "Stochastische Problemaufgaben" (Volume 23 of the IDM-series "Materialien und Studien", SCHOLZ 1981), the IDM is a supraregional central institute for research, development, advice, and coordination in the field of mathematical education. Using interdisciplinary and cooperative working structures and methods, the Institute is primarily devoting its activities to

- research and development related to the mathematics curriculum
- studies on the foundations of the didactics of mathematics as a discipline
- the investigation of processes and interactions in the mathematics classroom
- research and development related to the education and the continuing training of mathematics teachers.

I should like to add that, since its foundation in 1973, the Institute

has developed a national and international network for cooperation, coordination, documentation, and information. An international library and collection of materials, media, and documents has been established. Within these structures the IDM is holding conferences and workshops and offering a variety of services to individuals, groups, and institutions in the areas of research, development, practice, and administration. Furthermore, it is fundamentally concerned with the furtherance of young researchers in the didactics of mathematics by performing Ph.D. programs and organizing research seminars and colloquia.

For the sake of simplicity I shall group my following remarks according to the four above mentioned main R & D-components, starting with the curriculum.

As a structured body of knowledge, organized with respect to goals, contexts, learning activities, developmental aspects, a *mathematics curriculum* represents important features of societal reality and needs. Definitely, the problems related to the theme of this conference are of increasing importance for many domains in our society. More emphasis on the teaching of *probability and statistics* at school level seems to be an adequate consequence of this fact. However, it is questionable whether the usual way these topics are being presented is - if at all - sufficiently contributing to providing pupils with adequate knowledge, experience, and strategies needed for a good "decision making under uncertainty".

We have heard criticisms - at least related to German syllabi and textbooks (see SCHRAGE's paper) - that stochastics is often being taught from a predominantly mathematical viewpoint. Important methodological aspects such as alternative problem formulations, the process of modeling, competing models for a situation given, alternative decisions and their consequences, related risks and personal responsibilities, etc. are almost completely neglected.

G. v. HARTEN and H. STEINBRING in their paper indicate that important aspects of the genuine orientation of probability theory towards applications are deeply rooted already in the fundamental concepts of the theory. In particular, they explain this with respect to randomness and stochastic

independence. It is pointed out that these facts usually are not sufficiently elaborated in the standard approaches in school texts. The authors's suggestions are to get away with standardized problems in strictly fixed contexts, but rather develop systems of problems with flexible contexts which allow to develop a growing understanding of the interrelation between theoretical concepts and their meaning and role in various applications.

Clearly, the teaching of probability and statistics in the spirit of "decision making under uncertainty" asks for a *project type of activities* in school in order to guarantee a variety of important conditions: a sufficient degree of similarity with real situations, the possibilities for actually going into processes of development and modeling, the constitutive role of social factors and interactions. I want to emphasize in particular the opportunity which we find here to exhibit in a very specific sense the *social dimension of mathematics* and to give students concrete experience and methodological knowledge about this dimension and its various aspects such as: individual differences in viewing a situation and related problems, their influence on basic frames and decisions in the conceptualization of the situation and the formulation of problems, aspects of acceptability and agreement, personal or collective responsibility with respect to the situation given, especially to the human beings involved and their concerns, social handling of the co-existence of competing models, etc.

So far as first approximations to a *methodology of mathematization* which includes the social components have been developed (STEINER 1976), I see very interesting material for further elaboration of this methodology particularly in the paper by HUMPHREYS and BERKELEY on levels of knowledge representation in decision making but also in CROTT and ZUBER's paper on group shifts and polarization. Concerning the effect on the learner, I find of special interest SHAUGHNESSY's report on a problem and process oriented project-type course put into comparison with a traditional course in probability and statistics. When dealing with some of the (early) KAHNEMAN and TVERSKY tasks students in the small group activity-based course showed less reliance on representativeness and availability than students in the usual mathematics content oriented and teacher centered approach.

With respect to the *overall orientation of a contemporary formation* of our students through the study of mathematics and science at school, I think it is very important to not confine on traditional paradigms of teaching so-called value-free knowledge and follow a naive understanding of societal progress in applying this knowledge in technology and various domains of our modern life.

We should rather develop a critical attitude towards these paradigms and make students acquainted with more *comprehensive and integrative methods of thinking* (STEINER 1982). Systems theory and studies on decision making methodology and related human behavior as we find them in a variety of contributions to this conference are important sources from which we hope to receive substantial orientations for this difficult task.

These considerations bring us closer to the second working component which is related to the *foundations of didactics of mathematics* as a discipline. One of the central concepts in these foundations is the *concept of knowledge* - mathematical knowledge in particular - as a socio-historical reality. Thus epistemological, historical, sociological, and psychological studies of mathematical knowledge in its various contexts and representations play an important role in the work of the IDM (e.g., see BIEHLER 1982, JAHNKE 1978, OTTE & BROMME 1979, PFEIFFER 1981, STEINBRING 1980). Special studies have been devoted to the interrelations between mathematics as a field of organized development at the universities of the 19th century and mathematics as a topic of school curricula in the same period of time, especially from a philosophical and epistemological point of view (JAHNKE & OTTE 1981, BEKEMEIER et al. 1982).

Furthermore, the problem of *relating theory to practice* in teacher education and the teacher's professional work is a problem of various kinds of knowledge and their interrelations. First there is the problem of coordinating content related knowledge to pedagogical and psychological knowledge. Moreover, as has become evident in recent research on teaching, the teacher needs *knowledge about knowledge* (meta-knowledge) (see OTTE 1979, IDM-ARBEITSGRUPPE 1981). Investigations of these problems suggest the usage of various models, especially *data processing models*, approaches from an artificial

intelligence standpoint, models from PIAGETian psychology, etc. Here are close connections with some of the conference papers, the paper by HUMPHREYS and BERKELEY and the paper by R. GRONER, M. GRONER and F. BISCHOF on "The Role of Heuristics in Models of Decision" in particular. Also there are relations to investigations on *problem solving* as a central theme in mathematical education.

As for the *classroom* and the *learner* several conference papers are directly concerned with pupils' (and other subjects') difficulties, misconceptions, errors, and fallacies when dealing with statistical inference and more generally decision making under uncertainty (SCHRAGE, SHAUGHNESSY, GLAESER, BAR-HILLEL). GLAESER in his report on a doctor thesis investigating the correct and false ideas young pupils may have acquired with regard to certain phenomena of chance well before they began to study probability in the classroom claims that this type of study enables us to state today that mathematics education has become a science. In other papers the KAHNEMAN and TVERSKY findings, especially on base rate fallacies, play a central role. SHAUGHNESSY is presenting a variety of teaching strategies to improve the learners' understanding and behavior. Also many open research problems have been identified. This holds in particular with respect to the review by SCHOLZ and WALLER of the existing research on the probability concept in developmental psychology.

Of course, the *classroom* is a domain of *social interaction* and one is aware in the didactics of mathematics that meaning related to the content of instruction is in a fundamental way socially constituted (see BAUERSFELD 1982). The teacher is having a crucial role in this interaction which from a specific point of view has been described as a sequence of *establishing and breaking didactical contracts* between the teacher and individual or groups of pupils or also among different groups of pupils, depending on the social organization of the teaching-learning situation (BROUSSEAU 1982). I'm sure that social psychological models as have been discussed during this conference can be used to analyse these contract situations and their dynamics and that in this way professionally important knowledge for the *teacher as a decision maker* in the classroom can be gained.

This finally brings us to the fourth component of the above mentioned R & D activities: the *mathematics teacher*. We have already pointed out the importance of specific knowledge the teacher needs. It is evident that many of the conference topics stressed as being of value for the didactics of mathematics in general should also be considered with respect to their significance in the education of teachers. I believe that the publication of the conference report will induce processes in the community of mathematical educators which will take up the many impulses we received as participants.

References

- Bauersfeld, H. (ed.): 1982. Analysen zum Unterrichtshandeln. IDM-Untersuchungen zum Mathematikunterricht. Köln: Aulis
- Bekemeier, B. et al. (eds.): 1982. Wissenschaft und Bildung im frühen 19. Jahrhundert, Vol. I, II. IDM-Materialien und Studien Vol. 29 and 30. Bielefeld: Universität Bielefeld
- Biehler, R.: 1982. Explorative Datenanalyse - Eine Untersuchung aus der Perspektive einer deskriptiv-empirischen Wissenschaftstheorie. IDM-Materialien und Studien Vol. 24. Bielefeld: Universität Bielefeld
- Brousseau, G.: 1982. Tendances originales des recherches en didactique des mathématiques en France. Unpublished paper. IREM Bordeaux
- IDM-Arbeitsgruppe: 1981. Perspektiven für die Ausbildung des Mathematiklehrers. IDM-Untersuchungen zum Mathematikunterricht. Köln: Aulis
- Jahnke, H.N.: 1978. Zum Verhältnis von Wissensentwicklung und Begründung in der Mathematik - Beweisen als didaktisches Problem. IDM-Materialien und Studien Vol. 10. Bielefeld: Universität Bielefeld
- Jahnke, H.N. & Otte, M. (eds.): 1981. Epistemological and Social Problems of the Sciences in the Early Nineteenth Century. Holland: Dordrecht
- Otte, M.: 1979. The Education and Professional Life of Mathematics Teachers. In: B. Christiansen & H.G. Steiner (eds.), New Trends in Mathematics Teaching. Paris: UNESCO
- Otte, M. & Bromme, R.: 1978. Der Begriff und das Problem seiner Aneignung. In: J.R. Block et al. (ed.), Grundlagenkonzepte der Wissenschaftskritik als unterrichtsstrukturierende Momente. IPN-Arbeitsbericht Nr. 29. Kiel
- Pfeiffer, H.: 1981. Zur sozialen Organisation von Wissen im Mathematikunterricht. IDM-Materialien und Studien Vol. 21. Bielefeld: Universität Bielefeld
- Scholz, R.W.: 1981. Stochastische Problemaufgaben - Analysen aus didaktischer und psychologischer Perspektive. IDM-Materialien und Studien

Vol. 23. Bielefeld: Universität Bielefeld

Steinbring, H.: 1980. Zur Entwicklung des Wahrscheinlichkeitsbegriffs - Das Anwendungsproblem in der Wahrscheinlichkeitstheorie in didaktischer Sicht. IDM-Materialien und Studien Vol. 18. Bielefeld: Universität Bielefeld

Steiner, H.G.: 1976. Zur Methodik des mathematisierenden Unterrichts. In: W. Dörfler & R. Fischer, Anwendungsorientierte Mathematik in der Sekundarstufe II. Klagenfurt

Steiner, H.G.: 1982. Mathematisch-naturwissenschaftliche Bildung. IDM-Occasional Paper No. 22. Bielefeld: Universität Bielefeld

This Page Intentionally Left Blank

ADDRESSES

This Page Intentionally Left Blank

Gisela ALBERS, Dipl.Psych., Fakultät für Soziologie, Universität Bielefeld,
4800 Bielefeld, FRG

Wulf ALBERS, Prof. Dr., Institut für Mathematische Wirtschaftsforschung,
Universität Bielefeld, 4800 Bielefeld, FRG

Maya BAR-HILLEL, Dr., The Hebrew University of Jerusalem, Mount Scopus,
Jerusalem, Israel 91905, Israel

Dina BERKELEY, Dr., Decision Analysis Unit, The London School of Economics
and Political Science, Houghton Street, London WC2A 2AE, England

Walter F. BISCHOF, Psychologisches Institut, Laupenstr. 4, 3008 Bern,
Switzerland

Katrin BORCHERDING, Dr., SFB 24, Universität Mannheim, 6800 Mannheim,
L 13, 15-17, FRG

Helmut-W. CROTT, Prof. Dr., Psychologisches Institut, Albert-Ludwigs-Uni-
versität, Peterhof, 7800 Freiburg i.Br., FRG

Klaus FIEDLER, Dr., Fachbereich Psychologie, Universität Giessen,
6300 Giessen, FRG

Baruch FISCHHOFF, Prof., Decision Research, 1201 Oak Street, Eugene, Oregon
97401, USA

Georges GLAESER, Prof. Dr., IREM de Strasbourg, Université Louis Pasteur,
10, Rue du Général Zimmer, 67084 Strasbourg, France

Bernard GOITEIN, Institute for Social Research, University of Michigan,
Ann Arbor, Michigan 48109, USA

Marina GRONER, Dr., Institut für Psychologie der Universität Basel,
Bernoullistr., 4056 Basel, Switzerland

Rudolf GRONER, Prof. Dr., Psychologisches Institut, Laupenstr. 4, 3008 Bern,
Switzerland

Gerd v. HARTEN, Dr., Institut für Didaktik der Mathematik, Universität
Bielefeld, 4800 Bielefeld, FRG

Patrick HUMPHREYS, Prof., Decision Analysis Unit, The London School of Econo-
mics and Political Science, Houghton Street, London WC2A 2AE, England

Helmut JUNGERMANN, Prof. Dr., Psychologisches Institut, Technische Univer-
sität Berlin, Dovestr. 1-5, 1000 Berlin 10, FRG

Theo MAY, Dr., Abteilung für Experimentelle und Angewandte Psychologie, Univer-
sität Bielefeld, 4800 Bielefeld, FRG

Georg PFLUG, Prof. Dr., Fachbereich Mathematik, Universität Giessen,

6300 Giessen, FRG

Roland W. SCHOLZ, Dr., Institut für Didaktik der Mathematik, Universität
Bielefeld, 4800 Bielefeld, FRG

Georg SCHRAGE, Prof. Dr., Institut für Didaktik der Mathematik, Universität
Dortmund, Emil-Figge-Straße, Postfach 50 06 00, 4600 Dortmund 50, FRG

Ulrich SCHULZ, Prof. Dr., Abteilung für Experimentelle und Angewandte Psycho-
logie, Universität Bielefeld, 4800 Bielefeld, FRG

Reinhard SELTEN, Prof. Dr., Institut für Mathematische Wirtschaftsforschung,
Universität Bielefeld, 4800 Bielefeld, FRG

Zur SHAPIRA, School of Business Administration, The Hebrew University of
Jerusalem, Jerusalem Israel 91905, Israel

Michael SHAUGHNESSY, Prof. Dr., Department of Mathematical Sciences, College
of Science, San Diego State University, San Diego CA 92182-0314, USA

Heinz STEINBRING, Dr., Institut für Didaktik der Mathematik, Universität
Bielefeld, 4800 Bielefeld, FRG

Hans-Georg STEINER, Prof. Dr., Institut für Didaktik der Mathematik, Uni-
versität Bielefeld, 4800 Bielefeld, FRG

Werner H. TACK, Prof. Dr., Fachbereich Psychologie, Universität des Saar-
landes, 6600 Saarbrücken, FRG

Manfred WALLER, Prof. Dr., Abteilung für Experimentelle und Angewandte Psycho-
logie, Universität Bielefeld, 4800 Bielefeld, FRG

Thomas S. WALLSTEN, Prof. Dr., University of North Carolina, Chapel Hill,
Davie Hall 013A, Chapel Hill, N.C. 275 14, USA

Alf ZIMMER, Prof. Dr., Psychologisches Institut, Abt. Allgemeine Psycholo-
gie, Westf. Wilhelms-Universität Münster, Schlaunstr. 2, 4400 Münster,
FRG

Johannes A. ZUBER, Psychologisches Institut, Albert-Ludwigs-Universität,
Peterhof, 7800 Freiburg i.Br., FRG

ABSTRACTS

This Page Intentionally Left Blank

Wulf Albers & Gisela Albers: ON THE PROMINENCE STRUCTURE OF THE DECIMAL SYSTEM (pp. 271-287)

This paper gives a first approach to a theory of the prominence of numbers of the decimal system.

Basic components of the theory are the limited rational principles of rule construction by iterated addition or subtraction of a given amount, the refinement of a scale by adding the means of any two neighbours, and its coarsening by omitting all uneven elements.

Applying these principles to the powers of 10 (i.e. the most prominent numbers of the decimal system) various scales with different degrees of exactness can be constructed. Empirical observations suggest that among different scales with about the same exactness generally that one is preferred which can be constructed in the easier way, where the construction principles are those from above and the coarsening procedure is assumed to have the same complexity as the twofold refinement. This reduces the relevant scales to $(z \cdot a \cdot 10^n, z \in \mathbb{Z})$ with $n \in \mathbb{Z}$ and $a \in \{1, .5, .25 \text{ or } .2\}$.

The prominence of a number or a set of numbers is defined as the length of the steps of the coarsest scale containing the number or the set. (For empirical applications it is suggested to permit 25 percent exceptions.)

Empirical results indicate that the prominence of the set of numbers resulting from a specific decision situation is in many cases about one tenth of the numbers in question.

More precisely the prominence (or exactness of scale) of a specific decision problem seems to be selected according to the rule that it is maximal subject to the condition that the range of reasonable alternatives contains at least three prominent numbers.

Results of n-person bargaining experiments (with real valued characteristic function games) suggest that the evaluation of outcomes (more exactly; of pay-off differences) is done via the underlying prominence, such that the distance of two outcomes is evaluated by the number of prominence steps between them. By measuring incentives in this way a general theory of the probability that a "social" coalition (mainly an equal share coalition of all players) is formed could be developed for games of apex type, by which the empirical observation that the probabilities can be quite different if the same game is played on scales with different prominence can be explained. This result suggests that the prominence of the (spontaneously selected) scale can have basic importance for the analysis of individual or group decisions on numerical values.

Maya Bar-Hillel: THE BASE RATE FALLACY CONTROVERSY (pp.39-61)

The article attempts to sketch a conceptual and experimental history of the base rate issue. The review distinguishes between a social judgment paradigm and a textbook paradigm. Theoretical explanations of the base rate phenomenon, i.e. the representativeness heuristics, confusing or inverting conditional probabilities, the specificity factor, the causality factor, the vivid-

ness factor etc. are discussed with respect to these paradigms. On the basis of a report on various studies, the author hypothesizes that the different paradigms elicit different response tendencies, matching base rates on the one hand and judging representativeness on the other hand. She argues that different base rates may be justified in one and the same problem, depending on the causal structures accepted illustrating this by several examples. The consequences of this phenomenon for the role of normativeness and its (in-) determinacy are discussed.

(Abstracted by R.W. Scholz)

Helmut W. Crott & Johannes A. Zuber: BIASES IN GROUP DECISION MAKING: ANALYSIS OF GROUP SHIFTS AND POLARIZATION IN TERMS OF SOCIAL DECISION SCHEMES (pp. 229-252)

The present paper reports on three experiments about group decision making. The experiments deal with situations in which Ss with differing individual preferences have to reach a common decision. The results are first analysed according to the conventional methods of shift and polarization research. In all three experiments there are significant shift- and/or polarization effects. After the application of an appropriate Social Decision Scheme no systematic biases are left unaccounted for in two of the three experiments. The one exception is interpreted as a function of irregularities in preference orderings rather than a group bias. The results demonstrate that the decision schemes that take into account particularities in the obtained individual preference orderings are more effective in eliminating shift and polarization effects than those that are merely based on theoretical assumptions concerning the individual preference orderings.

Klaus Fiedler: ON THE TESTABILITY OF THE AVAILABILITY HEURISTIC (pp. 109-119)

TVERSKY and KAHNEMAN's availability heuristic, although originally intended to account for only frequency and probability judgments, has been used to explain almost all kinds of social judgments. Accordingly, the process of judgment formation is mediated by the availability of memorized information, that is, by the ease with which relevant material can be recalled at the time when the judgment is made. Recall operations - either pure retrieval or reconstructive recall - are regarded as the determining subprocess within the process of judgment formation. In this article, empirical evidence is presented which is hardly compatible with such an account. A reaction time experiment on the egocentric attribution phenomenon is described suggesting that Ss' claim to contribute more to various social activities than their partner is not caused by Ss' tendency to predominantly recall examples of their own activities. Within-judge correlations of recall and judgment latencies are rather low. And an analysis of the facilitating effect of prior judgments on subsequent recall latencies (for the same issues) does not reveal the kind of priming effect that would be expected if recall operations were already involved in the preceding judgments. Negative evidence from some other experiments on illusory correlations is also mentioned. These results are discussed in the context of methodological problems inherent in testing so-called judgmental heuristics.

Baruch Fischhoff/Bernard Goitein & Zur Shapira: SUBJECTIVE EXPECTED UTILITY: A MODEL OF DECISION-MAKING (pp. 183-207)

A constant concern of researchers who ask questions for a living is that respondents will be influenced by inadvertent aspects of the contexts in which their questions are set. Survey researchers, for example, have studied the effect of variables such as the race of the interviewer, the medium used (telephone, mail, in person) and the order in which questions are asked. When context effects are discovered, two approaches are possible. One is to treat them as research artifacts that need to be neutralized before questions can be meaningfully posed. The second approach is to treat context effects as facts of life that can provide useful insights into how people organize their beliefs and translate them into answers.

Both approaches may be found in the study of people's preferences. Each reflects and evokes a rather different picture of the nature of those preferences. In this talk, the two approaches will be contrasted, before turning attention to some recent attempts to develop theories of context. These theories predict powerful influences of seemingly irrelevant factors on the manner in which people express their preferences.

Georges Glaeser: THE PERCEPTION OF PROBABILITY SITUATIONS IN PUPILS AGED 12 TO 14 (pp. 313-323)

The probability concept is not simple: It can be dissolved into many components and we shall describe some of them.

From the genetical point of view, the student does not acquire all these components simultaneously, but rather crosses one threshold of the other. Hence, he/she always possesses but a partial understanding. He/she will be able to answer some questions which are compatible with the thresholds already attained; but he/she will systematically fail in other tests. Besides, the mere crossing of thresholds is not sufficient. Fixation is indispensable to secure the knowledge acquired. Otherwise, the pupil can be asked diverting questions which reveal the instability of the knowledge acquired.

This paper is a personal view of a study of Jesus ALARCON (Mexico), who presented his doctoral thesis in Strasbourg in Juni, 1982. This research is not in a genetical perspective, as this would require comparing the results of children from different age groups. The study used approx. 300 12- to 14-year-old pupils who were distributed to 3 samples of approx. equal size, i.e., 3 questionnaires were presented with slight alterations of the items to be tested. The results of only one of these samples (106 pupils) will be discussed here.

Rudolf Groner/Marina Groner & Walter F. Bischof: THE ROLE OF HEURISTICS IN MODELS OF DECISION (pp. 87-108)

In the initial chapter a review on the history of the field of Heuristics is given. In classical antiquity, two heuristic methods have been used, analysis and synthesis. Later, philosophers such as Descartes, Leibniz, and Bolzano introduced heuristic rules and systems, followed by mathematicians like Poincaré, Hadamard, and Polya. The history of heuristics is also closely connected to the development of algorithms.

In subsequent chapters, the role of heuristics in Psychology and Artificial Intelligence is briefly outlined. Four common characteristics of most approaches are proposed: (1) Representation in terms of information processing, (2) Restriction of the search space, (3) Logical sufficiency, and (4) Computational efficiency.

In the field of judgment and decision research, Herbert A. Simon's concepts of "bounded rationality" and "satisficing" provided a starting point for subsequent applications of heuristic methods. The heuristics of representativeness and availability by Tversky & Kahneman are reviewed and discussed, followed by some examples of information processing models in decision research.

Gerd v. Harten & Heinz Steinbring: RANDOMNESS AND STOCHASTIC INDEPENDENCE - ON THE RELATIONSHIP BETWEEN INTUITIVE NOTION AND MATHEMATICAL DEFINITION (pp. 363-373)

By means of historical investigations, epistemological reflections, and didactical analysis with respect to the notion of independence, we shall try to provide insights into the problem of a theoretical term and its applications. This will be the starting-point for stating some didactical theses about treating the notion of independence in the curriculum of Sekundarstufe I (lower secondary level) and will yield examples of their realization.

The difference between intuitive notion and mathematical definition reflects the insoluble tension between mathematics and reality. This should not be seen as a shortcoming, rather this tension has been one of the productive sources for the development of mathematics, and it ought to be the same for mathematics instruction.

Patrick Humphreys & Dina Berkeley: PROBLEM STRUCTURING CALCULI AND LEVELS OF KNOWLEDGE REPRESENTATION IN DECISION MAKING (pp. 121-157)

Much of the experimental work on decision making has been based on the assumption that a common understanding of the experimental task is shared by experimenter and subject. This assumption is challenged by the elaborations subjects have been found to impose in developing their understanding of the decision problem. Seven types of uncertainty are identified as in need of resolution in accounting for differences between these elaborations; one of them, procedural uncertainty, is seen as pointing to the need for the construction of a problem structuring calculus which will facilitate the under-

standing and the aiding of decision making. Previous attempts towards such a construction in terms of taxonomies of tasks by their analytic or substantive features are discussed and their drawbacks outlined. A multi-level problem structure generating scheme is proposed here and our first moves towards its definition are outlined. It is suggested that any decision problem is open to qualitatively different conceptualizations determined by the level of abstraction at which the decision maker is prepared and allowed to operate, given the support he or she gets from his or her environment in considering the problem. An outline of the characteristics of the levels is given; five levels of abstraction are identified and examples of typical properties of decision problems represented at each level are provided. The paper concludes with a discussion of the implications of this multi-level scheme for the study of individual decision making.

Helmut Jungermann: THE TWO CAMPS ON RATIONALITY (pp. 63-86)

The idea of rationality is still the core of decision theory, although the meaning of the concept has expanded considerably. In the last couple of years, a vehement debate has been led by two camps on the issue of rationality - or, in other terms, on the issue of biases, faults, and errors in human judgment and decision making. *One camp* claims that judgment and decision making under uncertainty often show systematic and serious flaws, indicating true deficiencies of the cognitive system. *The other camp* questions these conclusions and imputes them to the theoretical and empirical approach taken. After a description of the two camps and their implicit or explicit views on rationality, a resumé and outlook are given, focussing on the positive and productive aspects of the debate. The thesis and the antithesis seem to merge to a synthesis which includes in particular the recognition of the conditionality of normative models on assumptions about the environment and the cognitive representation of that environment.

Georg Ch. Pflug: DECISION-THEORETIC PARADOXES (pp. 375-383)

All methods for the estimation of unknown parameters or for testing of hypotheses require a twofold modelling process: One model for the underlying probability distributions and another model for the measurement of the efficiency of statistical procedures (loss function, risk function, etc.) is needed. Based on these models, which are usually found in intuitive way, the optimal statistical procedure can be derived in most cases. A paradox situation arises if this procedure does not coincide with the intuitively best or most adequate procedure for this situation. Some examples of that kind are collected in this paper. The question is: Which intuitively found model is not appropriate? Or should WALD's decision theory be revised?

Roland W. Scholz & Manfred Waller: CONCEPTUAL AND THEORETICAL ISSUES IN DEVELOPMENTAL RESEARCH ON THE ACQUISITION OF THE PROBABILITY CONCEPT (pp. 291-311)

This article provides a critical review of psychological theories and research approaches on the ontogenesis of the probability concept. An analysis of the conceptualization and interpretation of probability within developmental research reveals that with only one exception an objectivistic interpretation of probability has been made.

The reviewed (theoretical) research approaches, the cognitive developmental theory of PIAGET & INHELDER, FISCHBEIN's learning-developmental approach, and various information processing models differ in two main aspects. Firstly on the question whether the development should be considered to be a continuous or discontinuous process, and secondly the role of conceptual versus strategic knowledge for coping with probability problems is disputed.

The discussion tries to point out what progress could be gained by skipping a one-sided objectivistic interpretation of the probability concept and turning to a conceptualization of probability encompassing both sides, the objectivistic and subjectivistic view. This integration might also lead to a deeper understanding of the individual's conceptualization of uncertainty.

Georg Schrage: (MIS-) INTERPRETATION OF STOCHASTIC MODELS (pp. 351-361)

"Statistics is a body of methods for making wise decisions in the face of uncertainty" (W.A. Wallis and H.W. Roberts, *The Nature of Statistics*). In real applications these methods are sometimes misused by incompetent "experts". Some examples of this kind from different fields are presented. Misleading stochastic arguments can often be detected by pure common sense because of their consequences. Nevertheless, most people - including highly cultured people - accept even notorious nonsense if it is presented in scientific disguise. This fact indicates deficiencies in stochastic thinking. It should be a central aim of statistics education to make students sensitive to the abuse of instructive examples. An elaborated phenomenology of frequent errors including misleading strategies could be helpful to teachers at all educational levels.

Ulrich Schulz & Theo May: THE DEPENDANCE OF COGNITION AND BEHAVIOR IN 2x2 MATRICE GAMES (pp. 253-270)

A model for the iterated Prisoner's Dilemma Game and the iterated Chicken Game is proposed. The model considers beside choice behavior anticipation of the other player's choice and subjective certainty of this prediction. Information processing in the repeated game is due to possible changes in anticipation and subjective certainty. Choices depend on these variables.

By experimental variation of certain pay-off matrices it is demonstrated that some of the model parameters are task dependent.

Michael Shaughnessy: THE PSYCHOLOGY OF INFERENCE AND THE TEACHING OF PROBABILITY AND STATISTICS: TWO SIDES OF THE SAME COIN? (pp. 325-350)

In this paper we explore the interrelationships of research in judgment and decision making with research in mathematical education on the learning of probability concepts. The psychological literature demonstrates that people are subject to heuristics and biases when making inferences or probabilistic estimates. The literature of mathematics education indicates that many people are statistically illiterate. Thus, central motivating questions for the paper are: Can research in the learning and teaching of probability and statistics help the statistically naive judge? Can research in the psychology of inference help the naive statistician? How can research from both these disciplines aid the teacher of probability and statistics?

The paper consists of three main parts. Part one investigates obstacles to the use of statistics when making judgments or inferences. Part two discusses some suggestions from psychologists and from mathematics educators for increasing people's reliance upon statistics when making inferences. Suggestions cited for teaching statistics so as to improve people's inference and judgmental abilities include:

- 1) Teach from experiments to simulation to theory
- 2) Emphasize the confounding role of belief systems in judgment and the objectivity of statistics
- 3) Sensitize students to misuses of statistics in the media
- 4) Give students challenging problems to model by simulations
- 5) Expose students to Statistical Paradoxes
- 6) Expose teachers and students of probability and statistics to research results in the literature of judgment and decision making.

Specific examples, including sample problems, are given.

In part three, suggestions for further research are discussed. Among the questions raised and discussed in detail are:

- 1) How much of the reliance upon heuristics and biases is due to psychological factors and how much is due to inexperience with randomness and ignorance of statistical principles? Can instruction overcome people's tendency to ignore base rate data or rely upon heuristic principles? For example, what would be the effect of placing emphasis upon model building and simulation in instruction?
- 2) Is reliance upon causal schemas and deterministic reasoning a phenomenon of western culture, or could similar results be obtained in cross-cultural studies?
- 3) When groups make decisions, are they subject to the same heuristics and biases as individuals? If not, what are the implications for making judgments or for the teaching of statistics?
- 4) What is the role of the task itself upon the judgmental process?

It is suggested that cooperative research efforts between psychologists and mathematics educators be conducted in order to further investigate these questions.

Werner H. Tack: CONDITIONS OF VIOLATING INDIVIDUAL RATIONALITY (pp. 211-227)

In the context of conflicts of interest representable by n -person games in normal form, an outcome a is called *individually rational* if no player can enforce another outcome b which he prefers over a . The idea that any player can enforce certain outcomes by taking appropriate actions is based on the concept of control; different forms of control lead to different *control systems* represented as sets of restricted outcome sets. Since most empirical research on experimental cooperative side-payment games uses an experimental paradigm in which an isolated player can enforce only such outcomes which are of no value to him, violations of individual rationality are possible only in a very small fraction of all relevant experiments. Here - and in some experiments on non-side-payment games, too - their relative frequencies are in the magnitude order of 2 to 7 percent only. Despite this fact, they cannot be put aside as manifestations of otherwise unexplainable faults and errors, since there is some evidence that their probability depends on specific characteristics of the bargaining situation. More violations in situations with marked differences between the individually enforceable outcomes in connection with a high correlation between violations of individual rationality and equal share solutions lead to the hypothesis that they serve the function of diminishing existing interindividual differences.

Thomas S. Wallsten: THE THEORETICAL STATUS OF JUDGMENTAL HEURISTICS (pp. 21-37)

Past empirical research on judgmental heuristics and biases has focussed on questions of classification, with relatively little attention given to the development of general theoretical principles. It is the latter, however, that ultimately will lead to conclusions of greater generality and usefulness. A selected review of the literature on representativeness indicates some of the effects for which any complete theory must account and some of the limitations in many of the current experimental designs. The manner in which people use probabilistic information depends, in part, on (i) information specificity, (ii) information salience, (iii) individual differences, and (iv) problem wording. Theories are needed to elucidate these phenomena, and experimental designs are required that pay more attention to them.

Alf C. Zimmer: VERBAL VS. NUMERICAL PROCESSING OF SUBJECTIVE PROBABILITIES (pp. 159-182)

Ubiquitously people have to address the problem of uncertainty. The main focus of this article is an investigation of how people represent their knowledge about the uncertainty of events or regularities, and how they process this knowledge in order to make decisions in order to share their knowledge. In a first experiment it is asked how people gather information about frequencies of events and how this knowledge is interfered with the response mode, numerical vs. verbal estimates, they are required to use. The least interference occurs if the subjects are allowed to give verbal answers. From this it is concluded that processing knowledge about uncertainty by means of

verbal expression imposes less mental work load on the organism than does numerical processing. Possibility theory is used as a framework for modelling the individual usage of verbal categories. The 'elastic' constraints on the verbal expressions for every single subject are determined in a further experiment by means of sequential testing. The results from this experiment are used to suggest a simple mechanism underlying the "availability" heuristic. In further experiments it is shown that the superiority of the verbal processing of knowledge about uncertainty quite generally reduces persistent biases reported in the literature: conservatism and negligence of regression. In a final experiment about predictions in a real-life situation it turns out that in a numerical forecasting task subjects restricted themselves to those parts of their knowledge which are numerical. On the other hand subjects in a verbal forecasting task accessed verbally as well as numerically stated knowledge. The conjecture is made that the superiority of the verbal mode in representing and in processing knowledge about uncertainty is due to "shareability" (Freyd, in press) constraints which have evolved in the history of humankind and might even be phylogenetically determined.

This Page Intentionally Left Blank

AUTHOR INDEX

This Page Intentionally Left Blank

- Abelson, R.P. 344, 346
 Adams, J.L. 97, 105
 Agnew, N. 269
 Ainslie, G. 67, 83
 Ajzen, I. 45, 59, 328, 346
 Alarcon, J. 14, 314, 315, 319, 321, 322, 429
 Albers, G. 11, 17, 271, 287, 427
 Albers, W. 11, 17, 220, 226, 271 282, 287, 403, 427
 Allais 192
 Amarel, S. 88, 105, 152
 Anderson, J.R. 130, 152
 Anderson, N.H. 22, 34
 Anstadt, T. 226
 Aranoff, D. 270
 Arkes, H.R. 48, 59
 Arrow, K.J. 231, 251
 Asch, S.E. 22, 34
 Atkinson, J. 190, 204
 Aumann, R.J. 216, 219, 226
 Avis, N.E. 27, 36, 48, 60
 Baddeley, A.D. 188, 204
 Bar-Hillel, M. 8, 24, 26, 30, 31, 34, 39, 43-47, 50-52, 54, 59, 60, 122, 152, 327, 328, 343, 346, 347, 417, 427
 Barrios, A.A. 22, 34
 Barth, W. 255, 268
 Barthes, R. 145, 152
 Barton, C. 23, 37
 Bauersfeld, H. 417
 Beach, L.R. 27, 28, 35, 48, 59, 66, 71, 72, 83-85, 195, 203, 205, 207
 Begg, I. 170, 181
 Bekemeier, B. 416
 Bellows, N. 330, 348
 Benz, H.J. 344, 347
 Bergmann, A. 21
 Berkeley, D. 4, 5, 9, 12, 16, 71, 75, 76, 81, 83, 121, 124, 126-129, 151, 153, 308, 415, 417, 430
 Bernoulli, D. 121, 153, 159
 Berscheid, E. 404, 408
 Beyth-Marom, R. 48, 59
 Biehler, R. 418
 Birdsall, T.G. 123, 156
 Bischof, W.F. 9, 83, 102, 106, 417, 430
 Bixenstine, V.E. 255, 268
 Black, D. 231, 234, 241, 250, 251
 Bludszuweit, H. 255, 268
 Blundell, H. 255, 268
 Blyth, C.R. 380, 381
 Bogarth, R.S. 301, 308
 Bolzano, B. 93, 105
 Bonczek, R.H. 130, 139, 153, 430
 Borchering, K. 15, 403, 405, 407
 Borgida, E. 26, 34, 36, 39, 47, 59, 61, 328, 335, 348
 Boring, E.W. 95, 105
 Borkan, B. 200, 206
 Boxall, J. 150, 151, 156
 Brainerd, Ch.J. 300, 301, 303, 304, 309
 Braunstein, M.L. 203, 206
 Brehler, R. 416
 Brekke, N. 39, 59
 Bromley, D.B. 398
 Bromme, R. 5, 416, 418
 Bronner, F. 140, 153
 Brousseau, G. 417, 418
 Brown, A.L. 379, 399
 Brown, L. 381
 Brown, R.V. 129, 135, 136, 153
 Buckley, W.F. 339, 347
 Bühler, K. 95, 105
 Bunge, M. 134, 153
 Burnett, S.A. 129, 154
 Bushyhead, J.B. 25, 35
 Capon, N. 303, 309
 Caramazza, A. 165, 181
 Cardano 314
 Cardoze, S. 27, 36, 48, 60
 Carnap, N. 293
 Carroll, J.S. 26, 27, 35, 50, 51, 59, 203, 206
 Case, R. 304, 309
 Chaitin, G.J. 366, 373
 Chammah, M.A. 262, 269
 Chapman, J.P. 117-119
 Chapman, L.J. 117-119
 Chi, M.T.H. 33, 35
 Chomsky, N. 130, 153, 397, 399
 Choo D.T.G. 18
 Christensen-Szalanski, J.J.J. 25, 27, 28, 35, 48, 59, 72, 83
 Claparede, E. 98, 105
 Clark, A. 21
 Cohen, L.J. 22, 25, 35, 54-56, 59, 63, 83
 Cole, M. 399
 Coombs, C.H. 191, 204, 241, 251
 Corbin, R.M. 202, 206
 Corrigan, B. 188, 204
 Couturat, L. 89, 91, 105

- Cox, J.R. 33, 35
 Crandall, R. 26, 36, 47, 61, 328, 335, 348
 Crossmann, H.J. 287
 Crott, H.W. 11, 229, 234-236, 239 242, 251, 403, 405, 415, 428
 Cyert, R.M. 411, 412
 D'Alembert 314
 Davis, J.H. 231, 251
 Dawes, R.M. 188, 191, 203, 204, 204
 DeBono, E. 97, 105
 DeFinetti, B. 159, 181, 292, 309
 DeGroot, A. 96, 106
 DeHoog, R. 140, 153
 Delaney, H.D. 22, 35
 DeLoache, J.S. 399
 Dershowitz, A. 40, 59
 Descartes, R. 88, 89, 105, 430
 Deutsch, M. 255, 262, 268, 404, 407
 Dewey, J. 98, 106
 Dinges, H. 13, 16, 368, 373
 Dörner, D. 70, 83
 Doherty, M.E. 22, 36
 Doumouchel, W.H. 343, 347
 Dovalina, I. 27, 36, 48, 60
 Dubois, D. 165, 181
 Duncker, K. 96, 97, 106
 Duran, A.S. 34, 36
 Dusoïr, A.E. 29, 30, 35
 Eddy, D.M. 40, 44, 60
 Edgeworth, Y. 376-378, 381
 Edwards, W. 71, 83, 127, 135, 151, 153, 177, 178, 182, 192, 204
 Einhorn, H.J. 23, 35, 63, 71, 72, 76, 80-83, 104-106, 191, 194, 204, 205
 Ellsberg 192
 Elster, J. 73, 81, 83, 127, 153
 Emory, F. 287
 Estes, W.K. 49, 60
 Evans, J.St.B.T. 29, 30, 35, 80, 84
 Falk, R. 332, 337, 343, 347
 Feather, N.T. 79, 84, 184, 190, 203, 204
 Feigenbaum, E. 99, 106
 Feldman, J. 150, 154
 Feltovich, P.J. 33, 35
 Fermat 314, 357, 358
 Ferrell, W.R. 181
 Festinger, L. 237, 251
 Feynman, R.P. 367, 373
 Fiedler, K. 10, 23, 109, 428
 Fillenbaum, S. 21
 Fillmore, C.J. 130, 154
 Fine, T.L. 293, 309, 365, 373
 Fischbein, E. 299, 303, 304-306, 309, 331, 332, 345, 347, 353, 432
 Fischhoff, B. 10, 15, 27, 33, 35, 50-54, 58-60, 79, 84, 125, 135, 154, 155, 162, 171, 173, 181-183, 194-196, 199, 200, 205, 206, 326, 328, 347, 349, 387, 429
 Fitting, U. 269
 Flood, M.M. 295, 309
 Foa, E.B. 404, 407, 408
 Foa, U.G. 404, 407, 408
 Foltz, A.M. 40, 60
 Fong, G.T. 30, 36
 Ford, B. 335
 Forsyth, B. 21
 Freedman, D. 40, 60
 Freksa, C. 166, 181
 Freudenthal, H. 305, 309
 Freyd, J.J. 393, 394, 435
 Frindte, F. 255, 268
 Gabor, A. 287
 Gahagan, J.P. 255, 270
 Garey, M.R. 92, 93, 106
 Gasparski, W. 81, 84
 Gibson, J.J. 391, 394
 Gibson, R.O. 132, 155
 Ginosar, Z. 42, 60
 Ginsberg, R. 200, 206
 Ginzberg, E. 287
 Glaeser, G. 14, 313, 417, 42
 Glaser, R. 33, 35
 Goitein, B. 10, 79, 84, 183, 429
 Goldberg, L.R. 184, 187, 203
 Goldman, H.J. 262, 269
 Goodnow, J.J. 296, 309
 Goodwin, K.S. 299, 309
 Goulet, L.R. 299, 309
 Grange, J.W.J. 287
 Green, D.R. 309, 331, 332, 336, 345, 347, 353, 364, 373
 Gregory, R.L. 179, 181, 392, 394
 Grether, D.M. 200, 206
 Griggs, R.A. 33, 35
 Groner, M. 9, 87, 98, 102, 105, 106, 417, 430
 Groner, R. 9, 23, 87, 98, 102, 105, 106, 417, 430
 Gustafson, D.H. 140, 155
 Gyorgi, A. 148, 154
 Gödel, K. 92, 106
 Hacking, I. 159, 181, 292, 309

- Hadamard, J. 98, 106, 430
 Halpin, S.M. 255, 268
 Hankel, H. 88, 106
 Harten v., G. 15, 337, 363, 414, 430
 Hassebrauck, M. 404, 407
 Hassebrock, F. 34, 36
 Hausmann, L. 68, 84, 146, 155
 Hayes, J.R. 33, 35
 Heerboth, J. 68, 85
 Helmholtz v., H. 391, 394
 Nelson, H. 192, 204
 Heneman, H.G. 201, 206
 Herbrand 92
 Hersh, H.M. 165, 181
 Hesse, W. 255, 269
 Higgins, E.T. 112, 119
 Hintzman, D.L. 163, 181
 Hoemann, H.W. 303, 305-307, 309
 Hörz, H. 294, 310
 Hoffmann, P.J. 188, 204, 237, 251
 Hogarth, R.M. 4, 15, 23, 35, 63, 71-76, 80-84, 104, 106, 129, 154, 194, 205,
 Holmes, P. 336, 347
 Holsapple, C.W. 130, 139, 153
 Homans, G.C. 403, 407
 Hovland, C.I. 126, 154
 Howell, W.C. 129, 154
 Huber, O. 104-106
 Hubert, S. 22, 34
 Huff, D. 39, 60, 335, 338, 348
 Humphrey, G. 95, 106
 Humphreys, P.C. 4, 5, 9, 12, 16, 71, 75, 76, 81, 83, 121, 125-129, 137, 140, 141, 146, 148, 151, 153-155, 308, 415, 417, 430
 Illingworth, J. 269
 Inhelder, B. 13, 17, 291, 294, 297-300, 303, 305-307, 310, 432
 Irmscher, M. 163, 181
 Irwin, F.W. 22, 35, 36
 Isaac, D.J. 155
 Israel, R.B. 381
 Jacques, E. 134, 137, 147, 155
 Jahnke, H.N. 416, 418
 James, W. 381
 Janis, I.L. 65, 68, 69, 72, 82, 84, 202, 206
 Jaques, E. 132
 Jean, R.St. 405, 407
 Jepson, C. 29, 36
 Johnson, D.M. 123, 155
 Johnson, D.S. 92, 93, 106
 Johnson, E.J. 72, 84
 Johnson, E.M. 140, 155
 Johnson, P.E. 34, 36
 Johnson-Laird, P.N. 33, 37
 Jonas, K. 255, 261, 269
 Josephson, J. 296, 297, 310, 398
 Jsaac, D.J. 132
 Jungermann, H. 3, 9, 63, 68, 84, 146, 155, 431
 Kac, M. 364, 373
 Kahan, J.P. 220, 227
 Kahneman, D. 4, 8, 12, 17, 18, 22-24, 26, 28-32, 36, 37, 39-45, 49, 51, 53, 57, 59-61, 63, 65-68, 74-76, 78, 80-82, 84, 85, 103, 104, 107-109, 118, 119, 122, 123, 125, 147, 150, 152, 155-157, 170, 178, 179, 181, 192, 196, 198, 204-206, 306, 308, 310, 325, 327, 329, 332-334, 345, 348, 350, 382, 391, 394, 397, 398, 415, 417, 428, 430
 Kahr, A.S. 136, 153
 Kamen, J.M. 287
 Kassir, S.M. 39, 60
 Katz, N. 200, 206
 Kaufmann, A. 181
 Kaufmann, M.A. 165, 212, 213, 220, 226
 Kauke, M. 255, 268
 Kayser, E. 404, 407
 Keeney, R.L. 403, 407
 Kelley, H.H. 109, 113, 117, 119, 262, 268
 Kelsey, J.L. 40, 60
 Keynes, J.M. 373
 King, G. 112, 119
 Kircher, A. 90, 107
 Kistner, K. 405, 407
 Kleene 92
 Kleinmuntz, B. 105, 106
 Kleinmuntz, D.N. 105, 106
 Kneppreth, N.P. 140, 155
 Knoll, C. 234, 239, 251
 Kobasigawa, A. 304, 310
 Kohler, W. 96, 107
 Kogan, N. 405, 407
 Kohlberg, L. 397, 399
 Kolmogorow 314
 Konold, C. 329, 350
 Krantz, D.H. 29, 30, 36, 125, 155, 343, 347

- Kreuzig, H.W. 83
 Krivohlavy, J. 255, 269
 Kruskal, W.H. 348
 Kuhn, D. 303, 309
 Kuhn, T.S. 329, 348
 Kunreuther, H. 200, 201, 206
 Kyburg, H.E. 195, 205
 La Brecque, M. 334, 348
 Laing, J.D. 220, 226
 Lakatos, I. 203, 206
 Lamm, H. 230, 252
 Laplace 159, 292, 314
 Larichev, O.J. 154
 Lawrence, D.R. 237, 251
 Leach, E.R. 145, 155
 Leibniz, G.W. 88-91, 107, 430
 Leifer, R.P. 140, 155
 Lenat, D.B. 101, 102, 107
 Leventhal, G.S. 404, 407
 Levy-Bruhl 13
 Lichtenberger, J. 212, 219, 222, 226
 Lichtenstein, S. 27, 33, 35, 53, 58, 60, 135, 155, 162, 171, 173, 181, 182, 187, 188, 194-196, 199, 203, 205, 206, 326, 328, 334, 347, 349
 Lietzmann, W. 7, 17
 Lindley, D.V. 135, 153, 159, 182, 375, 381
 Lindman, H. 127, 135, 153
 Lingle, J.H. 117, 119
 Link, R.F. 348
 Livesley, W.J. 398
 Loomis, J.L. 255, 269
 Lopes, L.L. 392, 394
 Luce, R.D. 125, 155, 263, 264, 269
 Lullus, R. 88, 90, 93, 100, 107
 Lykken, D.T. 40, 60
 Lyon, D. 54, 60, 328, 348
 MacCrimmon, K.R. 406
 Mai, N. 269
 Manis, M. 27, 36, 48, 51, 52, 60
 Mann, L. 65, 68, 69, 72, 82, 84, 202, 206
 Manzat, J. 331, 347
 March, I.G. 411, 412
 March, J.G. 67, 74, 82, 84, 193, 204
 Markov 92
 May, T. 11, 174, 254-256, 259-262, 403, 432
 Mayer, A. 95, 107
 McAllister, D.W. 72, 84
 McCauley 44
 McClintock, C.G. 262, 269
 McCosh, A.M. 139, 155
 McCrimmon, K.R. 262, 269
 McFadden, W. 140, 154
 McGoey, D.J. 181
 McPhillamy, D.J. 199, 206
 Meehl, P.E. 40, 61
 Mery, J.L. 129, 154
 Messick, D.M. 262, 269, 406, 407
 Messik, S.J. 295
 Michaud, C. 129, 154
 Mikula, G. 404, 407
 Miller, D.W. 72, 85
 Miller, G.A. 11, 17, 187, 203
 Miller, L. 200, 206
 Miller, N.E. 199, 206
 Minsky, M. 102
 Mises v., R. 292, 310
 Mitchell, T.R. 71, 72, 83, 84, 193, 203, 205, 207
 Moller, J.H. 34, 36
 Monroe, K.B. 287
 Montgomery, H. 72, 85
 Morgenstern, O. 10, 356
 Morrison, R.J. 220, 226
 Moscovici, S. 230, 252
 Mosteller, F. 339, 348
 Müller, G.E. 96, 107
 Mulholland, J. 392, 394
 Murphy, A. 196, 206
 Myers, D.G. 230, 231, 252
 Mynatt, C.R. 22, 36
 Nelson, R.R. 411, 412
 Neumann v., J. 10, 356
 Newell, A. 94, 98-100, 107, 123, 156, 330, 348
 Ng K.O. 18
 Niessner, M. 226
 Nilsson, N.J. 101, 107
 Nisbett, R.E. 12, 17, 26, 29, 30, 34, 36, 47, 61, 63, 65, 85, 123, 156, 165, 166, 182, 326, 328-330, 333-335, 339, 343, 346, 348, 349, 391, 394
 Noelting, G. 304, 310
 Nozawa, G. 163, 181
 Oesterwalbesloh, K.W. 287
 Olian-Gottlieb, J.D. 201, 206
 Olson, C.L. 31, 36
 Orth, J. 95, 107
 Ostrom, T.M. 117, 119
 Otte, M. 416, 418
 Oxenfeldt, A.R. 287

- Pampu, J. 331, 347
 Parkes, E.M. 127, 156
 Pascal 314
 Payne, J.S. 203, 206
 Payne, J.W. 67, 72, 83, 85, 104, 105, 107
 Peevers, B.H. 398
 Peiters, R.S. 348
 Peleg, B. 216, 219, 226
 Peterson, C.R. 48, 66, 85, 136, 153, 195, 205
 Pfeffer, J. 193, 205
 Pfeiffer, H. 416, 418
 Pflug, G. 15, 375, 381, 431
 Phillips, L.D. 13, 17, 18, 63, 71, 75, 85, 123, 127, 129, 135, 136, 141, 150, 151, 153, 155, 156, 161, 162, 173, 177, 178, 181, 182, 195, 205
 Piaget, J. 13, 17, 131, 135-137, 156, 291, 294, 297-300, 303, 305-307, 310, 331, 397, 399, 400, 432
 Pilisuk, M. 255, 268
 Pisani, R. 40, 60
 Pitz, G.F. 68, 79, 85
 Platt, C.R. 200, 206
 Poincaré, H. 98, 107, 43
 Pollatsek, A. 329, 350
 Polya, G. 94, 97, 107, 108, 43
 Prade, H. 165, 181
 Purves, R. 40, 60
 Pylyshyn, Z. 255, 269, 391, 394
 Quinn, R.P. 196, 205
 Rade, L. 339, 349
 Raiffa, H. 403, 407
 Rapoport, A. 220, 227
 Rapoport, An. 262, 269
 Rasch, G. 174, 182
 Reed, H. 26, 36, 47, 61, 328, 335, 348
 Reinhardt, H.E. 342, 343, 349
 Reither, F. 83
 Rembde, M. 287
 Rescher, N. 393, 395
 Revenstorff, D. 255, 269
 Riker, W.H. 220, 227
 Rios, M. 150, 151, 157
 Rising, G.R. 348
 Roberts, H.W. 351, 43
 Ronen, J. 73, 85
 Rosen, A. 40, 61
 Rosen, L.D. 105, 108
 Ross, B. 303, 305, 307, 309, 310
 Ross, L. 12, 17, 63, 65, 85, 123, 156, 165, 182, 296, 297, 310, 326, 329, 334, 335, 343, 349, 391, 394, 398, 399
 Ross, M. 109, 117, 119, 349
 Rost, H. 13, 16
 Russo, J.E. 105, 108
 Sachs, N.J. 68, 85
 Sagi, P. 200, 206
 Salancik, G. 193, 205
 Sandler, A.M. 148, 156
 Sandler, J. 148, 156
 Sauer, C. 405, 407
 Sauermann, H. 11, 17, 287, 411, 412
 Savage, L.J. 127, 146, 156, 159, 176, 182, 192
 Scardamalia, M. 300, 303, 310
 Schaefer, H.J. 226
 Schaefer, R.E. 256, 269, 405, 408
 Schelling, T.C. 287
 Schermer, Th. 234, 239, 242, 251
 Schlenker, B.R. 262, 269
 Schmitz, G. 287
 Schoenfeld, A. 330, 349
 Scholz, R.W. 3, 5, 9, 11-13, 17, 21, 33, 36, 291, 344, 349, 375, 381, 413, 418, 428, 432
 Schrage, G. 14, 351, 361, 414, 417, 432
 Schütt, K.P. 176, 182
 Schulz, U. 11, 254-256, 259-262, 269, 403, 432
 Schwab, D.P. 201, 206
 Schwinger, T. 404, 407, 408
 Scott Morton, M.S. 139, 155
 Scribner, S. 399
 Secord, P.F. 398
 Selten, R. 11, 15, 17, 287, 409, 411, 412
 Selz, O. 95, 96, 108
 Sermat, V. 255, 262, 269, 270
 Shafer, G. 25, 36, 381
 Shaklee, H. 397, 399
 Shapira, Z. 10, 79, 84, 183, 429
 Shapiro, D. 33, 37
 Shaughnessy, J.M. 14, 325, 327, 333, 337, 345, 349, 353, 354, 360, 361, 415, 417, 433
 Shaw, J.C. 99, 107
 Shubik, M. 3, 17
 Shugan, S.M. 72, 85
 Shulte, A. 336, 349
 Sicoly, F. 109, 117, 119, 349
 Siegler, R.S. 26, 27, 35, 50, 51,

- 59, 301-304, 311
 Simon, H.A. 4, 11, 17, 33, 35, 36,
 65, 81, 85, 98, 99, 102, 103,
 107, 108, 123, 156, 187, 192,
 203, 287, 330, 348, 411, 412,
 430
 Six, U. 405, 408
 Sjöberg, L. 81, 85
 Slovic, P. 27, 33, 35, 53, 54, 58,
 60, 65, 66, 77, 84, 85, 123,
 155, 159, 171, 182, 187, 188,
 192, 194, 195, 199, 200,
 203-206, 326-328, 334, 346-349,
 391, 394, 397
 Smokler, H.E. 195, 205
 Snodgrass, J. 22, 36
 Solley, C.M. 295
 Solomon, L. 255, 270
 Sommer, L. 153
 Sowter, A.P. 287
 Snull, T.K. 112, 119
 Staudel, T. 83
 Staines, G.L. 196, 205
 Standel, F. 84
 Starr, M.K. 72, 85
 Staw, B.M. 201, 206
 Steele, M.W. 270
 Stein, C. 375, 378, 382
 Steinbring, H. 15, 337, 363, 414,
 416, 419, 430
 Steiner, H.G. 15, 413, 415, 416,
 419
 Stiegler, S. 382
 Stitt 44
 Stoner, J.A.F. 229, 230, 252, 408
 Suppes, P. 125, 155, 159, 182, 263,
 264, 269
 Svenson, O. 72, 85, 150, 154, 156,
 191, 203, 204
 Swets, J.A. 123, 156
 Tack, W.H. 11, 211-213, 219, 220,
 226, 403, 406, 434
 Tan, I. 18
 Tanner, W.P. 123, 156
 Taylor, S.E. 47, 48, 327, 349
 Tedeschi, J.T. 255, 270
 Terhune, K.W. 255, 270
 Thibault, J.W. 262, 268
 Thoene, B. 234, 239, 251
 Thompson, S.C. 47, 48, 109, 113,
 117, 119
 Thorngate, W. 123, 156
 Thraem, S. 226
 Tietz, R. 11, 17, 287
 Toda, M. 134, 146, 156
 Toman, R.J. 287
 Trope, Y. 42, 60
 Turing 92
 Turner, D. 336, 347
 Turoff, M. 12, 17
 Tversky, A. 4, 8, 11, 12, 17, 18,
 22-24, 26, 28-32, 36, 37, 39-45,
 49, 51, 53, 57, 59-61, 63,
 65-68, 74-76, 78, 80-82, 84, 85,
 103, 104, 107-109, 118, 119,
 122, 123, 125, 147, 150, 152,
 155-157, 159, 170, 178, 179,
 181, 182, 191, 192, 196, 198,
 204, 206, 306, 308, 325, 327,
 329, 332-334, 345, 348, 350,
 382, 391, 392, 394, 397, 398,
 415, 417, 428, 430
 Tweney, R.D. 22, 36
 Ulardt v., I. 68, 84, 146, 155
 Ulvila, J.W. 129, 153
 Vaihinger, H. 392, 395
 Vari, A. 127, 154, 157
 Vecsenyi, J. 127, 154, 157
 Venkatesan, M. 287
 Waerden van der, B.L. 94, 98, 108
 Wagner, C.H. 342, 343, 350
 Wald, A. 375, 376, 380, 382, 431
 Wallach, M.A. 405, 407
 Wallas, G. 98, 108
 Waller, M. 13, 15, 291, 397, 432
 Wallis, W.A. 351, 432
 Wallsten, G. 21
 Wallsten, H. 21
 Wallsten, T.S. 8, 9, 18, 21-23, 25,
 35, 37, 167, 434
 Wallster, G.W. 404, 408
 Walster, E. 404, 408
 Wason, P.C. 33, 37
 Watson, S.R. 148, 157
 Wegscheider, R. 269
 Weinstein, N.D. 190, 204
 Weir, M.W. 296, 306, 311
 Well, A.D. 329, 330, 345, 350
 Wertheimer, M. 96, 108
 Whalley, P.C. 18
 Whinston, A.B. 130, 139, 153
 White, S.E. 193, 205
 Wickelgren, W.A. 97, 108
 Wilks, S.S. 188, 204
 Wilson, T.D. 166, 182, 330, 348
 Winkler, R.L. 196, 206, 256, 270
 Winter, H. 352, 361
 Winter, S.G. 411, 412

- Winterfeldt v., D. 129, 130, 148,
151, 157
Wisudha, A. 18, 137, 154
Witte, W. 166, 167, 182
Wittgenstein, L. 182
Wolff, G. 7, 18
Wooler, S. 141, 155
Wright, G.N. 13, 17, 161, 182
Wright, P. 188, 204
Wyer, R.S. 112, 119, 196, 205
Zadeh, L.A. 161, 182, 393, 395
Zavalloni, M. 230, 252
Zimmer, A.C. 10, 15, 135, 157, 159,
165, 166, 170, 181, 182, 391,
393, 434
Zuber, J.A. 11, 234, 242, 251, 403,
405, 415, 428
Zwicky, F. 94, 108

This Page Intentionally Left Blank