Proceedings

of a

Research Conference

on

Subjective Probability

and Related Fields

April 10-12, 1969

at the Psychological Institute
of the University of Hamburg

Proceedings of a

Research Conference on Subjective Probability and Related Fields

held from April 10 - 12, 1969, at the Psychological Institute of the University of Hamburg

Contents :

Editor's Preface

This note is mainly for those who did not attend the conference to tell them what all this is about. Carl-Axel Stael von Holstein, Charles Vlek, and me had the feeling that it might be good to have such a conference to join all people working somewhere - and mostly rather isolated - on subjective probability and related fields. We decided to have this conference in Hamburg since we have here quite a bunch of students working in this area, and since it seemed to be located most centrally (considering participants from Sweden and Italy, Poland and England) - but the main part of the job of mailing circulars, and organizing the program, was done by Charles Vlek from Leiden.

It was, actually, rather hard to get together a list of prospective participants who might be invited to such a conference - and I am afraid we never managed to complete this list. So, our apologies to everybody we omitted, and our plea to tell us everybody' s name who should be on the list, so we can invite him (or her) next time.

Yes, there will be a next time - probably in spring 1970, to be organized this time by Gerard de Zeeuw, H.C. van der Meer, Charles Vlek and W.A. Wagenaar. Everybody on the "list of prospective participants" will be invited to that next conference, and everybody interested in it but not on the list should contact Gerard de Zeeuw to get onto it.

The sessions of this year's Research Conference on Subjective Probability were mainly devoted to unpublished research and unsolved problems - since published papers are easier to be read at home. The following papers are just summaries of the participants' contributions and discussions, in the order they gave them during the Conference - with one exception : I put Prof. de Finetti's paper to the end since it contains some comments which are important to all of the preceding ones.

Thanks to everybody who helped make this Conference and the edition of these Proceedings possible !

Dirk Wendt

General Overview of Current European Research in the Area of Subjective Probability and related Problems

Dirk Wendt (Hamburg)

In accordance with the general aim of this Conference, my overview will be confined to studies under way, or accomplished recently but unpublished, as far as I know of them. This restriction will give my review a rather strong bias in favour of research done at Hamburg which I have to apologize for - but it is clear that these studies are best known to me.

Trying to be systematic, I will first mention some experiments in the context of the SEU model, then some on risk preferences and the utility-of-gambling problem, and finally some on subjective probability (SP) scaling, probability revision, and conservatism.

In the field of the SEU model, Günter Lehmann is (at least since 1964) attacking the problem of simultaneous measurement of SP and utility. He will talk himself about his recent research. Matthias Burisch tried to find out whether there exists a special utility of gambling. He had his Ss scale the utilitiesfor various quantities of two different commodities (cigarettes and candy) by means of the Marschak bidding procedure, and then had them evaluate gambles with certain amounts of cigarettes as oneoutcome, and the S's equivalent in candy as the other outcome. Results showed that many Ss evaluated the gambles higher than the riskless offers of the same commodities (indicating a positiv utility of gambling), some of them showed no utility of gambling at all, and one S had a negative one.

Heiner Imkamp compared two methods of utility scaling, the Marschak biddingprocedure, and pair comparisons. His Ss were customers of a Stuttgart department store who evaluated various commodities by the two methods. Imkamp, too, had his Ss scale the utilities of gambles involving two different commodities as outcomes.

Miss H.C. van der Meer has contributed several experiments to the problem of risk preferences-- see her list of references--,and, in Hamburg, Bernd Stein, RoswithaKoch, Rainer Kakuska, and me have worked in this field, starting from Coombs'theory of an ideal level of risk (Portfolio Theory). Kakuska gives a separate report in these Proceedings. Stein, Koch, and me are testing the hypothesis that a certain perceived level of risk may be attained either by varying the range of a gamble, or by varying the uncertainty about the probability of winning or losing. These two determiners of risk should work compensatorily so that an S should increase the range of a gamble chosen when the uncertainty about the probability involved decreases.

Somewhere between choices among risky alternatives (bets) involving both SP and utilities, and straight SP research, I have to mention some studies on "subjective levels or confidence". This concept is derived from the normative model of classical hypothesis testing (i.e. Neyman-Pearson), and applied to S's behavior in experimental situations. Jürgen Kriz has criticized this concept as not descriptive of S's sensation of uncertainty in a decision between competing hypotheses. Helmut Jungermann has compared the classical hypothesis testing model to the Bayesian model in an experimental setup, and Harm Stehr found out that it is not at all unimportant whether you have an S estimate the probability of the rejected, or of the accepted hypothesis: in the latter case, conservatism is considerably larger.

Studies devoted to SP directly (without connection to values in bets) can be divided into two areas : those concerned with prior probabilities, and those in a probability revision context where Bayes Theorem is considered as normative model. In the first field, Carl-Axel Stael von Holstein has done some experiments (see his report), and so has Reiner Fricke. Fricke studied the influence of context on probability (or, rather, relative frequency) estimates in a sequence. His Ss had a kind of multiple probability learning task. He found that the mean objective relative frequency worked as a kind of adaptation level on the estimates. Another multiple probability learning experiment has been done by Wolfgang Manz. He varied the relative frequency of the most frequent event, and the distribution of frequencies over the other events, and got almost correct estimates under all conditions. Hermann Rüppell is working on an experiment in probability learning with shifting objective event frequencies over the series.

Concerning the conservatism phenomenon in probability revisions, people in and from Michigan have developed three hypotheses: misperception, misaggregation, and experimental artefact (see Edwards 1968: Conservation in Human Information Processing, in: Kleinmuntz, Formal Representation of Human Judgment. Wiley, New York, p. 17-52). An experiment by Diethard Freitag was directed towards an evaluation of the first two of these hypotheses. He gave his Ss a long training period to get familiar with the data sources (bookbags with red and green beads) untill they gave satisfactorily unbiassed estimates of P (D|H), and then had them decide between competing hypotheses on the contents of a bookbag, based on information from a sampling process with optional stopping. Compared to optimal stopping calculated from objective, and from estimated P(D H), Ss bought too large samples, thus favoring the misaggregation hypothesis.

Experiments aimed at the influence of various variables on conservation in the bookbag-and-pokerchip paradigm have been performed by Manz in Köln., by Vlek in Leyden., by Kriz in Vienna, by Genser in Hamburg, and by Grabitz in Mannheim. As far as I can see, most of their results favored the experimental artefact hypothesis. Wolfgang Manz studied the effect of sample size (n = s + f) and sample composition (s/n) on conservatism. He essentially concluded that conservatism might be an experimental artefact since in daily life s/n is more important than s-f. Jürgen Kriz, in another experimental setup, came to similar conclusions. He, and Charles Vlek, and Reiner Fricke, and me, too ,tried to find some general function which relates SP to objective probabilities(OP), an attempt which has been criticized by Prof. de Finetti (see his contribution in this Proceedings). Gerd Gekeler studied the effect of sample size on such SP/OP functions. Burkhard Genser investigated the relative influences of objektive posterior odds and sample probability on conservatism. Dirk Revenstorff investigated relationships between OP and SP in a dice game. Ss had to estimate the probability to roll a higher score than a given score. They overestimated low, and underestimated high probabilities.

Two major points of criticism have arisen against all these types of experiments: First, whether it makes sense at all to search for a function relating SP to OP. Prof. de Finetti will make some remarks on this topic, see his contribution in these Proceedings. Second, whether the bookbag-and-pokerchip paradigm is not too artificial at all for normal Ss such that it leads to lots of experimental artefacts which have no relevance on actual behavior in daily life. My personal impression from these discussions is that we should "bury that urn" and abandon the bookbag-and-pokerchip paradigm as a tool of research completely.

My apologies to everyone I omitted, misunderstood or misinterpreted in this "general overview" -- as I told you, I tried to give you some impressions on unpublished studies, and as such, my report was mainly based on private communications, and gossip.. All kinds of additions and comments are welcome.

Some Problems in Revision-of-Opinion Studies

Charles Vlek (Leiden)

One of the things one would like to be able to do with Bayes' theorem is handling the following situation. Suppose one wants to predict which of several political parties a particular person or group of persons is going to vote for. The set $\{H_i\}$ consists of the existing parties or the subset of parties that are worth considering. The set $\{D_j\}$ of available data describes the person or the group whose voting behavior is to be predicted. Setting up a matrix with the H_i as colums entries and the D_j (which we suppose to be independent) as row entries, we shall have to determine for each cell the value of $p(D_j/H_i)$. An additional row of the matrix would specify the prior probabilities of the various hypotheses. Once we have succeeded in determining the cell entries of the matrix we shall be able to compute

$$p\ (H_i/D_1, D_2, \ldots, D_j, \ldots, D_k)$$

for each column, if we have processed k data sequentially. The rationale behind such an experiment would be the assumption that people are generally unable to aggregate lots of single pieces of information, but that they are capable of estimating the diagnostic impact of single data.

In the past 5 years one has begun attacking this kind of problem by reducing it to the well-known two-hypotheses bookbag-and-pokerchip situation or "urn paradigm". We shall confine ourselves to the probability-revision-version of the urn-paradigm, the other version being the information-purchasing or deferred decision making situation. Let the bags contain red and blue chips, be symmetrically composed, and let p_r be the proportion of red chips in the predominantly red bag. r represents the numberof red chips in the samples, which contains a total of n chips.

Many variables seem to influence the height of the subjective posterior probability (SPP) given after S has been presented with two bags and a handful of chips. A common finding is that SPP increases as a function of 2r-n, but Ss are more conservative for higher values of 2r-n. A well-established fact seems to be that SPP decreases when n increases (Peterson et al., 1965; Pitz, 1967; Vlek, 1965; Vlek & Van der Heijden, 1967), although Manz (1968) found the opposite result. Equally well-established seems to be that SPP is rather independent of p_r (Vlek, 1965; Pitz et al., 1967; Marks, 1968), which of course causesthe Accuracy Ratio to be a decreasing function of p_r, as Phillips & Edwards (1966) observed. The latter authors also reported that SPP is dependent upon the pay-off scheme and the type of judgment scale used (direct SPP-,odds-, and log odds estimates lead to different result).

In order to explain the generally observed conservatism effect two hypotheses have been advanced. One (Petersen et al.,1968) says that Ss properly appreciate the diagnostic impact of a single datum, but that they aggregate these impacts in a nonoptimal way when presented with a sequence of data. The other hypothesis (Vlek, 1965; Beach, 1968) says that people misperceive the diagnostic impact $p(D_j/H_i)$ of a datum. The issue has not yet been settled unambiguously, since there seems to be some confusion about whatshould be called "a" datum (one chip, or asequence of chips). It seems justified, however, to conclude that conservative $p(H_i/D_j)$'s are at least partially due to "conservative" $p(D_j/H_i)$'s, i.e., subjective sampling distributions (the pattern of $p(D_j/H_i)$-estimates for all D_j under a given H_i) are flatter than the corresponding objective ones (Ptersen et al., 1968; Wheeler & Beach, 1968 Vlek & Van der Heijden, 1967). A recent, seemingly irrational finding (Manz, 1968) is that SPP's increase with the value of r/n in the sample.

On the basis of these various proofs of irrationality we may join Pitz et al. (1967) in concluding that " a subject's performance in a probability revision task is nonoptimal in a more fundamental way than is implied by discussions of conversatism. Performance is determined in large part by task characteristics which are irrelevant to the normative model."

Recognizing the importance of subjective sampling distributions (SSD's) as as subjective expectation patterns, Vlek & Van der Heijden (1969) recently attempted to infer SSD's from direct posterior probability estimates obtained in a two-hypotheses bivariate normal distribution experiment conducted by Lichtenstein and Feeney (1968). They argued that every SPP, when converted into log posterior odds and given equal prior probabilities, would give them the difference between the logs of the ordinates of the two overlapping distributions. Multiple regression analyses on 3 different transformations of the stimulus characteristics showed that the inferred sampling distributions leading to the best "prediction" of the actually given SPP, were far deviant from the normal distributions supposed to have been learned by the subjects. In addition, subject widely differed in the weights they attached to the stimulus information. In an extended and somewhat modified (as yet unpublished) version of Lichtenstein and Feeney's experiment it was observed that many Ss subdivide the probability scale into as few as only 5 categories.

On the basis of the above-mentioned difficulties, some of which seem to be inherent to the experimental situation, and some of which are due to an improper understanding of what is needed before Ss can meaningfully express their subjective probability, the following problems-of-further-research may belistet.

1. The study of subjective probability - because of its preference for objectively specifiable chance situations - is in danger of proceeding along experiments that have a very low external validity (if not also a ry low internal validity). Care should be taken to design representative experiments.

2. Instead of having to rely on such introspectionistic measures as subjective probability estimates, one would like to use behavioral measures which are monotonic with variations in subjective probability.

3. Because posterior probabilities are considered as the result of an information processing activety, the information to be processed be unambiguously defined (e.g. is S processing absolute or relative numbers of chips, or likelihoods, or likelihood ratios?).

4. If S is supposed to "make use of" sampling distributions one should distinguish between the extent to which S has learned the objective sampling distributions, and the rule he follows, he follows in converting his $p(D_j/H_i)$'s into a $p(H_i/D_j)$. An evaluation of the extent to which people are " Bayesian" can only be meaningful, if we know exactly which $p(D_j/H_i)$'s S has "in mind".

5. The apparant functioning of SSD's necessitates the study of multiple probability learning, especially in situations where $p(x_i)$ has a systematic functional relationsship to x_i.

6. The measurement of learned subjective probability distributions urges the development of simple, understandable methods for assessing these distributions. Winkler (1967, a and b, 1968) and Toda (1968) give starts to this problem.

7. In complex situations it might be inevitable to treat probability estimates as if they convey only ordinal information.

References:

Beach, L.R. J. exp. Psychol., 1968, 77,57-63

Lichtenstein, Sarah, & Feeney, G.J. Org. Behav. Hum.Perf., 1968, 3,62-67

Manz, W. Unpublished 'Abschlussbericht' Univ.Cologne, August 1968

Marks, D.F. Bull. Brit. Psychol. Soc., 1968, 21, 115(Abstract).

Phililips, L.D. & Edwards, W. J. exp. Psychol.,1966,72,346-354

Petersen, C.R., Schneider, R.J. & Miller, A.J. J.exp. Psychol., 1965, 69, 522-527

Peterson, C.R., Du Charme, W.M. & Edwards, W.J.exp.Psychol., 1968, 76, 236-243

Pitz,G.F. Psychonomic Science, 1967, 8, 257-258

Pitz,G.F., Downing, L. & Reinhold, Helen Canad.J.Psychol, 1967, 21, 381-393

Toda, M. in: Proc. NUFFIC Intern. Summer Session "Algebraic Models in Psychology", Psychol. Inst. Univ. Leiden, November 1968

Vlek,Ch.Report 009-65 Psychol.Inst.Univ.Leiden, December 1965

Vlek,C.A.J. & Van der Heijden,L.H.C. Report E 017-67 Psychol.Inst.Univ. Leiden, Nov.1967

Vlek,C.A.J. & Van der Heijden, L.H.C. Org.Behav.Hum.Perf. 1969, 4, 43-55

Wheeler, Gloria, & Beach, L.R. Org. Behav.Hum.Perf., 1968, 3, 36-46

Winkler, R.L. J. Amer.Statist. Assoc.,1967, 62,776-800

Winkler, R.L. J. Amer.Statist. Assoc.,1967, 62,1105-1120

Winkler, R.L. Management Science, 1968, 15, B61-B75

Strategies in Probability Learning.

Wolfgang Manz (Brighton, Sussex)

A recurring finding in probability learning experiments is the phenomenon of "probability-matching" (PM). In these experiments Ss are to choose repeatedly between two alternative responses A and B under a noncontingent reinforcement schedule in which A is rewarded with probability p and B with probability (1-p). Ss' response frequencies averaged over a run of consecutive trials can be described as occuring with probabilities π and (1-π). PM is said to result when π asymptotically approaches p. It is not only a well documented finding (LUCE and SUPPES, 1965 for a summary), but can also be derived from the assumptions of mathematical learning theories (BUSH and MOSTELLER, 1955; ESTES, 1959). Outside the realm of learning theories PM has acquired somewhat of the status of an explanatory concept suited to be incorporated as a basic premise into theories of decision-making (GULLAHORN and GULLAHORN, 1963; KEMENY, SNELL, and THOMPSON, 1966; HALPIN and PILISUK, 1967). How sound, however, are generalizations based upon PM?

Criticism has been levelled against the psychological meaningfulness of the basic assumptions in mathematical learning theories, and thus the soundness of the derivation of PM may be questioned (FOPPA, 1964). This is of minor importance here as the empirical findings cannot be negated. The most succinct criticism of the empirical findings is that of TODA (1964) who argues that there is not only lacking in most experiments an appropiate statistical test (cf. GRANT, 1962), but that the generally adopted procedure of averaging over trials and Ss is bound to obscure rather than to elucidate any regularities in Ss' behaviour. Finally, a few experiments have failed to produce PM both in the two-choice situation (EDWARDS, 1961; NEIMARK and SHUFORD, 1959) and especially in situations involving more than two choices (GARDNER, 1958, 1961; GARDNER and FORSYTHE, 1961; COTTON and RECHTSCHAFFEN, 1958).

Seen in the context of this conference two interpretations based on the above mentioned findings are of special interest.

The first pertains to the view which takes the asymptotic response probabilities π as a measure of the S's subjective probability of p. PM in that interpretation is then seen as indicative of the S's ability to estimate correctly the objective probability p only in the case of a two-choice situation. With more than two choices overestimation of the more frequent and underestimation of the less frequent events seem to obtain. The first inference as to the S's being correct in estimating binomial probabilities in a learning task, however, contradicts the findings of over- and underestimating in comparable judgmental tasks as reprted by STEVENS and GALANTER(1957) and SHUFORD (1961). This leads to the problem of reconciliating the conflicting evidence (CUBE, 1965). The answer to this problem is to reveal it as a pseudoproblem: the behavioural index of choice proportions π and $(1-\pi)$ reflects only in a limited way the S's estimate of the unknown probabilities p and (1-p) which govern the stimulus presentation. This cannot be taken as a straightforward measure of subjective probability because the response based on the subjective estimate is always mediated by strategical considerations on the part of the S. Any explanation of the PM behaviour therefore has to look at this aspect of the problem. There is experimental evidence demonstrating that the estimate of the objective probabilities and the choice behaviour as reflected in the response probabilities are two different irrelated aspects of the situations which should not be confounded (NEIMARK and SHUFORD, 1959; MANZ, 1968).

The second interpretation sees the experimental paradigm of a probability learning task as a game against nature in which nature is known to be in either of two states A or B with fixed but unknown probabilities p and (1-p) and where the player-subject has to select a strategy p $(\pi,(1-\pi))$ with which to predict the two states. The pay-off matrix for most of the experiments with positive reinforcement attached to the correct predictions and no reinforcements to the incorrect predictions can be represented as follows:

		NATURE	
		state A	state B
PLAYER	A	1	0
	B	0	1

From this it immediately follows that for any $p \neq (1-p)$ only a pure strategy P (π=1.o, (1-π)=0.o) is a rational solution. Any mixed strategy like the PM strategy is clearly suboptimal. Why then do Ss prefer the suboptimal strategy as seems to be well documented in the experiments?

There have been several solutions suggested to this problem, the most superficial one being the assertion of Ss' basic irrationality. More sensible solutions differ among themselves but nevertheless point in the same direction: that of underestimating the impact of the "demanded characteristic" (ORNE, 1959), of the whole experimental situation. SIEGEL (1964) introduces the negative utility of boredom operating in these simple experiments as explanation for the Ss' suboptimal choices. In his experiments he varies the pay-off matrix to include losses, that is negative reinforcement for incorrect predictions, or likewise introduces complications in the procedure to reduce boredom. His Ss stop then using PM as the most preferred strategy and tend to adopt a pure strategy instead. Others have drawn attention to the run structure of the sequence of trials which can give Ss wrong but compelling clues and thus produce recency effects (GOODNOW, 1955; JARVIK, 1951). Variations of situational context are equally effective in reducing PM (HOKANSON and DOEHR, 1964; MANDLER, COWAN, and GOLD, 1964). Elaboration of the implications of the experimental situation reduces Ss' suspicion of being tested for their capacity to detect a systematic pattern behind an effective disguise of a simple experiment. Thus the appropiate framing of the experimental set-up leads to a well-defined task and bears out the same result of Ss' adopting the pure strategy (NEIMARK and SHUFORD, 1958; MANZ, 1968).

All this highlights the basic design fault in most of the probability learning experiments: the staightforward transfer of an experimental situation designed to study rat behaviour at a choice point (BRUNSWICK, 1939) into a paradigm for doing research into human decision making has all the virtues of simplicity and elegant control of the stimulus but is bound to challenge the Ss' capacities to redefine the experimental situation for themselves thereby rendering all control to uselessness. As a pointed summary statement one is tempted to say that PM as result in these experiments is more indicative of loose thinking in the experimenter's part than it is revealing anything about Ss' behaviour. The question of how Ss use the probabilistic information accumulated over the consecutive trials for developping strategies how they execute, test and refute them or stick to them is still open for experimentation.

References:

Bush, R.R. and F. Mosteller: Stochastic models for learning. New York: Wiley and Sons, 1955

Brunswik, E.: Probability as determiner of rat behaviour. J. Experim. Psychol. 1939, 25, 175-197

Cotton, J.W. and Rechtschaffen, A.: A replication report: Two- and three-choice verbal-conditioning phenomena. Experimental Psychol., 1958,56, 96

Cube, F. v.: Kybernetische Grundlagen des Lehrens und Lernens. Stuttgart. Klett 1965

Edwards,W.: Probability learning in 1ooo trials. J. Experm. Psychol. 1961, 62, 385-394

Estes, W.K.: The statistical approach to learning theory. In: Koch, S. (ed): Psychology: A study of Science, Study I, Vol 2, 38o-294. New York, Toronto, London, 1959

Foppa, K.: Probabilistische Lernmodelle. In: Bergius, R, (ed) Handbuch der Psychologie, 1. Band, Allgemeine Psychologie.I. Der Aufbau des Erkennens. 2. Halbband: Lernen und Denken. Seite 617-64o. Göttingen 1964

Gardner, R.A.: Multiple-choice decision behavior. Amer. J. Psychol. 1958, 71, 71o-717

Gardner, R.A.: Multiple-choice decision behavior with dummy choices. Amer. J. Psychol. 1961, 74, 2o5-214

Gardner, R.A.and Forsythe, J.B.: Supplementary report: Two-choices decision behavior with many alternative events. J. experim. Psychol. 1961, 62, 631

Goodnow, J.J.: Determinants of choice distributions in two-choice probability situations. Amer. J. Psychol. 1955, 68, 1o6-116

Grant, D.A.: Testing the Null Hypothesis and the strategies and tactics of investigating theoretical models. Psychol. Review 1962, 69, 54-61

Gullahorn, J.T. and Gullahorn, J.E.: A computer model of elementary social behavior. Behav. Science 1963, 8, 354-362

Halpin, S.M. and Pilisuk, M.: Probability matching in the Prisoner's Dilemma. Psychol. Science, 1967,7, 269-27o

Hokanson, J.E. and Doerr, H.O.: Probability learning of interpersonal events. J. Person. 1964, 32, 514-53o

Jarvik, M.E.: Probability learning and a negative recency effect in the serial anticipation of alternative symbols. J. Experim. Psychol. 1951, 41, 291-297

Kemeny, J.G., Snell, J.L., Thompson, G.L.: Introduction to finite mathematics. New York: Prentice-Hall 1966 (sec. edit.)

Luce, R. and Suppes, P.: Preference, utility, and subjective probability. In: R. Luce, R.R. Bush, E. Galanter (eds.), Handbook of mathematical psychology, vol. III, p. 249-41o. New York: Wiley & Sons 1965

Mandler, G., Cowan, P.A., Gold, C.: Concept learning and probability matching. J. Experim. Psychol. 1964, 67, 514-522

Manz, W.: Der Einfluß der Reizverteilung auf die Strategiewahl beim Wahrscheinlichkeitslernen. Psychol. Forsch. 1968, 32, 169-184

Neimark, E.D. and Shuford, E.H.: Comparison of predictions and estimates in a probability learning situation. J. Experim. Psychol. 1959, 57, 294-298

Orne, M.T.: The nature of hypnosis: artifact and essence. J. Abn. Soc. Psychol. 1959, 58, 277-299

Shuford, E.H.: Percentage estimation of proportion as a function of element type, exposure time, and task. J. Experim. Psychol. 1961, 61, 43c-436

Siegel, S.: Choice, strategy, and utility. New York:Mc Graw Hill 1964

Stevens, S.S. and Galanter, E.H.: Ratio scales and category scales for a dozen perceptual continua. J. Experim. Psychol. 1957, 54, 377-411

Toda, M.: The micro-structure of guess processes. State College. Pennsylvania: Report No. 4, Devision of Mathematical Psychology, Institute for Research 1964

Non-random Behaviour in Subjective Randomisation Tasks.

W.A. Wagenaar (Soesterberg, Netherlands)

Many experiments are dealing with the subjective concept of "chance". One often encountered method of studying this topic is asking the subject to produce a long randomized sequence of alternatives out of a certain choice-set. The sequential qualities of a real random series are determined by chance only, and the supposition in randomisation experiments is that sequences randomized by Ss will show something of what Ss expect that will happen by chance.

About 15 post-war publications describe this type of experiments but these experiments differ with respect to a number of factors.

1. the number of alternatives varies from 2-26.
2. the nature of alternatives varies among letters, digits, nonsense syllables, heads/tails, push-buttons, circles on a paper etc.
3. the experimental situation requires different modes of production: calling out, writing down, pointing, stamping, pushing on buttons.
4. the length of the sequences varies from 20 to over 2500.
5. the number of previous responses visible varies from no to all responses.
6. the rate of response varies from 0.25 to 4.0 sec. per response.
7. production can be paced or unpaced.
8. the number of Ss varies from 2 to 125.

There are not two experiments which have all conditions exept one in common, therefore difference in results can never be traced to one single factor.

The same confusion exists with respect to the meaning of randomness, as defined by the measures for non-randomness, actually used. A great variety of zero and first-order measures was used, and only in a few cases second and higher order measures.

As may be expected, the results of these experiments are very often in contradiction with each other. As an example three experiments on the influence of rate of production were compared. Baddeley (1962, 1966) found that increase of rate gives increase of non-randomness. His hypothesis was that the amount of information a subject can generate per second is limited. Teraoka (1963) found decrease of non-randomness with increasing speed, whereas Warren & Morin (1965) found a small increase. The latter authors found that the information generated per second is increasing with rate. Some objections against these three experiments are:

1. No external representation of the choice set. In case of many alternatives, the experiments are dealing with activation of the choice-set rather than with selection out of the choice set.
2. No indication of the nature of the deviation from randomness (positive or negative recency).
3. Small numbers of Ss in two experiments.
4. Low order of analysis.

In an experiment carried out in our Institute we used:

- 2,3,4,6 and 8 alternatives (push-buttons)
- sequnces with 1oo elements.
- 8 ss
- rates of 0.5, 1, 2, and 4 sec./response.
- analysis up to sixth-order dependencies.

we found:

- no effect of rate.
- dependencies significant to the sixth-order values.
- large individual differencies for lower order dependencies.
- small individual differencies for higher order dependencies.
- negative recency for the higher order dependencies.

The main research problems are:

1. experimental situation: task, instruction, apparatus?
2. measurement of non-randomness: which measure, to which order of analysis?
3. relevance: are the results representative for the subjective concept of chance?
4. model: is there any model describing the behaviour of all Ss, and is there any theory to explain why Ss are bad randomizers?

Some Problem in the Practical Application of Bayesian Decision Theory

Carl-Axel S. Staël von Holstein (Stockholm)

In almost all practical applications of Bayesian methods it is assumed that there exists a prior distribution and that it is known. Only seldom is there any discussion of the problem of how to assess these subjective probability distributions. The choice of assessment technique is not obvious and experiments have shown that subjects react in different ways when confronted with different techniques. It may also be so that a technique which is good in one situation may not work well in another situation.

There is not only a need for different techniques for the assessment of subjective probability distributions, but also for empirical evidence of how well they work in practice. Empirical research will lead to improved and more reliable techniques.

The paper is intended to give a review of some problems in connection with the assessment of subjective probability distribution for practical applications. These problems include

1. The choice of assessment technique or techniques.
2. The meaning of "good probability assessors".
3. Scoring rules for the evaluation of assessments.
4. The aggregation of several persons' assessments.

The various problems will partly be treated against the background of results from empirical research.

The paper also includes a review of some practical applications.

1. The choice of assessment technique or techniques

The chapter reviews a number of assessment techniques. The following four techniques do not depend on any special assumption about the analytic form of the subjective probability distribution.

(i) Direct assessment of fractiles. This could be done by means of either successive subdivisions or direct questions regarding fractiles.

(ii) Consideration of the effect of sample evidence on the subject's probabilities.

(iii) Expression of prior judgments in the form of equivalent prior samples. This means that the subject should determine two numbers r and n such that his knowledge would be roughly equivalent to having observed r "successes" in n trials.

(iv) Assessment of points on the probability density function (by means of questions regarding relative densities and relative areas).

The method of successive subdivisions refers to the following scheme. Let p be the unknown quantity of interest. We can then proceed to ask the assessor in the following way:

1. Determine a point $p_{.5}$ such that it is equally likely that p is less than $p_{.5}$ as that p is greater than $p_{.5}$.
2. Assume now that you were told p in fact is less than $p_{.5}$. Conditional on this information determine a new point $p_{.25}$ such that it is equally likely that p is less than $p_{.25}$ as that p is greater than $p_{.25}$ (i.e. is between $p_{.25}$ and $p_{.5}$).
   (A similar question is asked for the case that p is greater than $p_{.5}$. This gives a third point $p_{.75}$.)
3. One can continue to ask the assessor to subdivide a given interval into two equally likely parts. This should be continued until there are enough fractiles to give an understandin of the shape of the distribution function.

The chapter also discusses such assessment techniques which have been designed for situations where it is assumed that the subjective probability distribution is a member of some family, e.g. the family of beta distributions.

Finally some empirical results are presented. They show that different techniques often imply different distributions especially if the assessor has not had a good statistical training. This makes it necessary to use more than one technique to check the assessments against each other.

2. The meaning of "good probability assessors"

It seems natural that there are many reasons why one could be interested in measuring the goodness, whatever that may mean, of a probability assessment. This could be done in order to compare several assessors for a selection of one or more of them for further assessments. Another reason could be that one wants to measure the effect of some kind of training, i.e. if the assessments have become "better" after a period of training.

An assessor can be good in essentially two respects. He may have some knowledge of probability concepts and he makes assessments that are consistent with the theory of probability. The assessments should not only obey the postulates of this theory but they should also correspond to his true judgments. The second respect concerns the assessor's knowledge of the practical problem at hand. Assume that a person is to predict the buying price of some share after a given period. It is then desirable that he has good knowledge about the various factors that may influence the price of the share in question. These two standards of goodness can be called the normative and the substantive standard of goodness, respectively. To summarize, the normative standard of 'goodness' concerns expertise in probability assessment, while the substantive standard of 'goodness' concerns expertise in the domain in which the assessments are made.

3. Scoring rules for the evaluation of assessments.

A scoring rule is a measure of the goodness of an assessment. You may, for example, ask a person to assess a distribution for the temperature at noon the following day. This temperature can be observed and the person's score will be a function of the assessment distribution and the actual value of the temperature. The chapter contains a discussion of various desiderata for scoring rules and several examples are presented.

Scoring rules can be used for the following purposes. We assume that we have found a scoring rule which constitutes a valid measure of goodness in a certain situation. We can then use this scoring rule to rank any set of assessors and the scores can serve as a basic for the selection of assessors for future assessments. The scores can also be used to determine weights for the assessors if their assessments are to be aggregated in some fashion. The last use to be mentioned here concerns the training of assessors. The scores represent an evaluation of the assessor's performance and an examination of them may give indications as to how he might improve his assessments. He can learn to better understand the correspondence between judgments and probabilities and thus become a better assessor in the normative sense. The scores will also be useful for interpersonal comparisons. A person with consistently low scores can compare his assessments with those made by successful assessors. He may find, for example, that he consistently overestimates the quantity studied or that the confidence expressed in his assessments are not justified by the outcomes.

4. The aggregation of several person's assessments.

So far we have been concerned only with individual assessments of SPDs. There are, however, many practical situations where one might have assessments made by several peaple. How should one proceed to make use of these SPDs which represent the judgments of the assessors and which, of course, may be quite different? The answer is to aggregate the SPDs into one distribution, which can be used as the basis for a decision.

A simple way of aggregating distributions is to take a linear combination of them. This requires that the assessors are assigned weights (w_i). Let $F_i(x)$ be the i'th person's distribution function and let $F^x(x)$ by $F^x(x)=\sum w_i F_i(x)$. The difficulty in this method lies in the choice of an appropriate set of weights. If there is no information available on the experience or the capability of the assessors, then one would probably choose to assign equal weights to them. It seems natural, however, that the various assessments should carry unequal weights when there is some information available on the past performance of the assessors.

The chapter reviews some suggestions for the choice of weights. It also contains descriptions of other ways to arrive at a consensus of several assessors.

The consistency of subjective probability judgments

David Marks (Sheffield)

It has been shown by de Finetti (1937) that the necessary and sufficient conditions for a person to avoid having a book made against him is that his subjective probabilities (SP's) for a complete set of incompatible events sum to one. This condition is known as "consistency" or "coherence". Three experiments are reported which investigate the extent to which the SP's of real people, rather than ideal ones, conform to the mathematical norm of consistency.

1) Using the well-known urns situation, each S was asked to give an SP estimate that a sample S_i was taken from Urn 1 and, on another trial, that S_i was taken from Urn 2. Fourteen out of 34 S's adopted the simple strategy: SP=r/n, which of course yielded perfect consistency. Fifteen of the remaining S's gave SP's which summed to slightly, but significantly, above one. This effect was caused by greater conservatism for low objective probabilities (OP's).

2) From the finding that conservatism is greater for low OP's it can be predicted that total SP will increase as the number of alternative events increases. This prediction was verified using situations involving up to 4 mutually exclusive events. Total SP averaged at 1.14 for a 4-urn situation.

3) A rather different approach studied the consistency of degree of belief given for propositions of the type "A man will be landed on the moon by November 1970". SP's were obtained for several pairs of contradictory statements, one statement in each pair being positive, and one negative. SP's summed to greater than one for statement-pairs with a moderately probable positive statement, and to less than one for statement-pairs with moderately improbable positive statement. This effect was created by greater conservatism for negative statements.

These experiments suggest the following conclusions: (a) although biases are present in SP judgments, there are regularities in these biases of a simple kind; (b) it is important to take account of such biases in making use of SP for inferences and decisions.

Reference: De Finetti, B.: "La prévision: ses lois logiques, ses sources subjectives", Annales de l'Institut Henri Poincaré, 1 (1937)

Experiments on Information Purchasing

Bernt Larsson (Lund)

Introduction: "I am not quite sure of what is meant by experiments on information purchasing. Broadly speaking, it must be all experiments where the subject has the possibility to get more information before making his final decision. However, this definition includes most experiments performed to-day in the Bayesian area and is therefore not very useful. One way to restrict the set of relevant experiments is to look at the dependent variables of main interest. For the following discussion I will select the efficiency of the subject's behaviour and, coupled with this, the number of observations purchased and the strategy for the final decision. I will come back to a definition of efficiency but let us first discuss some aspects of choosing an experiment."

The choice of an experiment: "The kind of situation which I have in mind is the very common one where the subject should state from which population his sample of observations has been taken. Correct choice of the population gives a zero loss but a wrong choice often involves a great loss in relation to the cost of one observation." Several questions are put and discussed in order to choose an experiment: 1) From how many populations can the subject simultaneously take samples? 2)How many populations are there to choose between when making the final decision for a sample? 3) Which data processing model can be used - if any? 4) Is sampling fixed or sequential? 5) Has the subject full knowledge of the parameters defining the situation (Shuford's concept of a well-defined situation)? 6) Are the situations hypothetical or does the subject actually lose something during the experiment? 7) Is the subject told if his final decision is right or wrong? 8) Are the independent variables between-subjects variables or within-subjects variables?

The concept of efficiency: Efficiency is defined as the ratio between two total losses (loss from choosing a population plus coat of observation). "The total loss can be of two kinds: the actual amount lost in a situation or the expected amount an infinite number of samples or replications of a trial. In general, the two total losses defining an efficiency ratio may come from different sources. You can have two models, two subjects, a subject and a model, a subject and a group of subjects, and so on. So the concept is rather general and can be used for different comparisons." The concept is then illustrated for the case of fixed binomial sampling for simple hypothesis testing. "I have mostly dealt with a dependent variable called efficiency, which may be useful when describing individual decision making. If calculations of efficiency are based on actual losses and comparisons are between subjects it is not necessary to use a statistical model but for all other cases you must have one, as far as I can see. The model has given two things here: an optimal behaviour and a possibility to calculate the average amount lost by the subject over an infinite number of replications of the situation. The last thing is only valid if the subject's choices of the sample size and the critical value when choosing a hypothesis are not systematically changed over replications. It is only experiments which can verify or deny this assumptions But if it is true the model shows how to use observed behaviour in a special trial to estimate average efficiency over many trials not performed."

Footnote: The concept of efficiency for Bayesian decision making will be discussed in a forthcoming report.

Subjective probabilities and utilities in non-verbal decision-processes.

Günter Lehmann (Göttingen)

The purpose of this investigation was the simultaneous measurement of subjective probabilities and utilities which are assumed to determine non-verbal decision processes between uncertain events.

An experiment was made, in which a subject had to make series of subsequent partially reinforced decisions t (t=1,5o) between two alternatives a_1, a_2 (keys to be pressed). After each decision t for a key a_k he could win with a constant probability p_k, known only to the experimentator, a constant amount v_k of "points" (v=1,2o), the sum of which could be changed into money (pennies) after the experiment. 24o of such decisions series, each of them having another combination of constant outcomes v_1, v_2 with constant probabilities p_1, p_2, were made by 157 subjects, whose decisions and gains were automatically enregistrated.

Since the true probabilities p_1, p_2 of winning v_1, v_2 were not known to a subject his decisions could be interpreted as paired comparisons between "sample exspected values" $E_{tk} = p_{tk} \cdot v_k$ with the sample probabilities:

$$p_{tk} = \frac{\text{number of reinforcements at key } a_k \text{ after decision } t}{\text{number of choices of key } a_k \text{ after decision } t}$$

which varied stochastically in each series: $\lim st\ p_{tk} \to p_k (t=1,\ldots)$

By dividing the continua of probability P and of "points" V into 1o equally spaced classes $p_i (i=1,1o)$ and v_j $(j=1,1o)$ the experimentator could distinguish 1oo different sample expected values $E_{ij} = p_i . v_j$. The decisions (over 1o.ooo) of all series were combined in a paired comparison matrix A, in which a_{ijkl} represented the number of preferences of E_{ij} over E_{kl}.

I. The determination of subjective expected values:

Applying the Luce-axiom I :

(1) $$p_{ijkl} = \frac{a_{ijkl}}{a_{ijkl} + a_{klij}} = \frac{W_{ij}}{W_{ij} + W_{kl}}$$

there was calculated a matrix W of subjective expected values from A (Ford-algorithme): $w_{ij} \Leftrightarrow E_{ij}$

Results:

1. The L_{max}-value of the maximum-likelihood-solution for W from A was significantly high ($p < .o5$).

2. The prediction of decisions by using the E_{ij}-values was highly significant ($p < .ooo1$). But the gain in the number of correctly predicted decisions by using w_{ij} instead of E_{ij} was signif. ($p < .o5$) (randomization test).

3. The first Eigenvectors of the matrices WW' and W'W, giving the least-squares-solutions for the subjective continua S and U of probability and " value of points", had a significantly high variance of 91.7%. This was in favour of the hypothesis that w_{ij} is composed by the product of the two subjective values s_i and u_j.

4. The variation of the rows and columns of W and the continua S and U showed a) an overestimation of low probabilities, b) an antiogive form of the utilities, i.e. the well-known subjective similarity between the numbers ("points") from

"1o" to about " 14".

II. The comparison of the hypotheses of multiplicative resp. additive connection (conjoint measurement problem):

$$w^{*}_{ij} = s^{*}_{i} \cdot v^{*}_{j} \quad \text{versus} \quad w^{+}_{ij} = s^{+}_{i} + v^{+}_{j}$$

In order to determine more directly the subjective continua S and U, expression (1) was decomposed by introducing a subjective operator o

$$(2) \quad p_{ijkl} = \frac{s^{o}_{i} \circ u^{o}_{j}}{s^{o}_{i} \circ u^{o}_{j} + s^{o}_{k} \circ u^{o}_{l}}$$; o is multiplication or addition resp.

The two maximum-likelihood-solutions for the continua S^{*},U^{*} resp. S^{+},U^{+} were approximated from A.

Results:

1. The multiplicative continua S^{*} and U^{*} predicted the decisions significantly better ($p < .o2$) than the additive continua S^{+} and U^{+}. (randomiz tion test).
2. The same tendency held for the rank-correlation between W and the least-squares-solutions: $\sum (w_{ij} - s^{o}_{i} \circ u^{o}_{j})^2 \rightarrow \min.$ ($o \equiv \cdot$ resp. $+$).
3. The multiplicative and the additive continua were similar to one another and to those determined from W in section I.

An experiment on risk preferences

Rainer Kakuska (Hamburg)

The experiment reported purports to test the so-called 'portfolio theory' by C.COOMBS. This theory tries to explain the attractiveness of bets by the expected value (EV) and the risk involved in playing a bet. The EV is computed in the usual way, but no attempt is made to derive subjective measures of the payoffs and the probabilities of winning and losing. The measure of risk is not specified by COOMBS; in this experiment the 'expected loss' (probability of losing x amount lost) was selected. The theory assumes the existence of a 'point of ideal risk' on every EV-level. The bet having this ideal amount of risk should be preferred to bets having the same EV but higher or lower risk. The ideal risk is supposed to increase with increasing EV.

Procedure: A simple type of two-outcome bet is used, where the probabilities of winning and losing sum up to 1.oo and the payoffs can be set arbitrarily. This is realized by a device comparable to a wheel of fortune with a rotating spinner on a disc divided in two sectors; the sizes of the sectors determine the probabilities. Bets on 5 different EV-levels are used, with 5 different amounts of risk on each level. This is done for three distributions of probabilities separately to avoid possible influence of probability preferences. After the subject has acquired some familiarity with the type of bet by actual playing the bets are presented graphically in random order and are rated by the subject according to their attractiveness. The procedure is repeated several times in order to get stable measures, but care is taken not to bore the subject too much. For the same reason, i.e. motivation, a proportion of the bets is actually played and the subject is payed his gains. To make sure that the subject doesn't give his ratings at random, the bets with the highest attractiveness ratings have the greatest chance of being selected for playing.

Processing of the data will be done separately for each subject and will be mostly plotting curves of equal attractiveness (indifference curves) in the EV-risk space. The assumptions of the theory predict a certain form of indifference curves which can be compared to the form of the curves obtained empirically.

Variing types of random sequences and relative efficiency of strategies.

Gernot Kleiter (Salzburg)

An outline of some psychological experiments.

Considering <u>one class</u> of random events <u>uncertainty</u> for Ss is not always given in the same way. <u>Different types</u> of distributions could be distinguished, uniform distributions and normal distributions for example. During one session <u>shifts</u> may occure for different parameters (range, mean, standard deviation) and type of distribution. Events in the considered class must not be independent, so, kind and degree of <u>dependency</u> could be varied also.

So (i) kind of distribution, (ii) shifts, and (iii) degree of dependency are the main independent variables to be considered in this experiment.

Strategies of Ss in handling those different characteristics may be different also. Most often, however, it is difficult to identify "subjective" strategies because within one session different strategies are applied. It seems reasonable to get an over-all measurement for the efficiency for all strategies applied as compared with (i) a "constant" random strategy and (ii) with a constant optimal strategy. This does not exclude the possibility to identify subjective strategies for this experimental procedure.
So (i) the relative efficiency of varring "subjective" strategies and (ii) - if identification is possible - the strategies applied by the Ss are the main dependent variables.

Procedure

The efficiency of S's strategies in handling sequences of random events from one population will be investigated in a man-machine-interaction situation similar to a gambling situation.
S sits in front of a desk computer. The computer has been programmed to give an output of random numbers (x_i) with different parameters possible.
Instruction is given like this:
"Suppose you want to sell bananas every week. You have to buy your supply on Monday; but you don't know how many people will buy bananas every week. There are many random influences. All you know is: certainly 1 banana and not more than 100 bananas will be bought; between these limits every demand has an equal chance (for uniform distribution). The computer tells you the demand which is entirely independent of your supply. For every sold banana you will win 1 unit (r units), for every banana you couldn't sell your loss will be the same (c units). So for every banana lying one week on your stockpile you lose 1 unit for bananas are perishable indeed. Your task is to try to find the supply for every week (trial) which handles best the random demand and to get out a maximum mean win over a long period of weeks".
One trial is given by:
S input of his supply (S_i)
computers output of random demand (x_i, for uniform distribution for example between 1 and 100)
and additionally S's win in this trial (actual win this week, g)
and S's mean win over all trials until know (g_i)
One session: 200-300 trials.
At the end of each session the mean win can be payed to the Ss.

General ypothesis: Ss are doing better in all situations where they are confronted with distributions with a marked central tendency. From real life they have got more experience with these problems. So Ss are expected to make higher relative mean wins in managing normal distributions than uniform distributions, normal distributions with small standard deviations are better handled than normal distributions with large standard deviation etc.. Ss are expected doing better with a symmetric outcome function ($r = c$). Ss are expected furthermore to do not very much better than a random strategy for they comprehend a sequence of random numbers not as independent events. They usually try to make a prediction for the next random number with respect to the last 3 or 4 random numbers and hope to get a perfect hit (strategy of getting "maximal" - not optimal win by trying to "solve the problem").

Theoretical notes

This type of game is defined by:

(i) constants: r: win for every sold unit
c: loss for every unit not sold
W: range of demand when uniform distribution and
$\bar{x}$, s : mean demand and standard deviation of demand when normal distribution.

(ii) output : random demand x_i

(iii) input : supply S_i

Optimal strategy for uniform distribution is given by:

Constant input (!) $S_{opt.} = \frac{2rW + r + c}{2(r + c)}$

The expected mean win with this strategy is:

$E(\bar{g})\, S_i = \frac{S_i}{2W} \left(2rW - (r+c)(S_i - 1)\right)$ where $S_i = S_{opt.}$

In the discussion, the author was recommended:

(a) to use a more dynamic paradigm,

(b) to get information from real market policies of supply

(c) to get a clear-cut comprehension of what is meant by 'random' in this device.

The TRUE SP PROBLEM (Subjective Probability Problem) versus various more or less related SIDE PROBLEMS

Bruno de Finetti (Roma)

1.- The point to be stressed

The comments I consider appropriate now, after listening to the reports and discussions of this Conference, may be summarized by the recommendation of a clear distinction between the true SP Problem and the various more or less related side-problems.

The true SP problem consists in the investigations concerning the ways in which probabilities are assessed and used by more or less educated people, and the way in which such abilities may be improved. This seems to me the field in which the cooperation between all specialists concerned is most wanted, and that is particularly true for the expected contributions from the psychologists.

Side-problems are not necessarily less important, when of interest for particular reasons or topics, but, in order to avoid misleading confusions, they ought not to be interpreted as properly belonging to the true SP problem.

The point to be stressed is not really different from that I, since before, intended to illustrate here in an intervention: that is, the distinction between the unitary view ("all probabilities are subjective"), and the dualistic - or maybe pluralistic - ones ("there are several kinds of probabilities: objective and subjective"). But my task seems now much easier, because I can avoid a good deal of abstract philosophical discourses by seizing the opportunity of remarks concerned with some topics from the present discussions.

2.- An unexpected terminological ambiguity

I must begin, indeed, trying to settle an unexpected preliminary question: maybe a merely terminological one, but likely to create inextricable misunderstandings. I was surprised to see that some psychologists call subjective probability any "wrong" evaluation of the "correct" probability that is called objective. Without questioning here whether and how far such notions could be regarded as meaningful, it must be warned that, in the established terminology, "subjective probability" has a completely different meaning. It is the degree of belief as evaluated by anybody making the best use of all the information available to him, and of his own skill.

The distinction just mentioned, between the unitary and the dualistic (or maybe pluralistic) point of view, will further clarify the issues on the adjectives "objective" and "subjective". Objectivists call "objective" such probabilities that they accept, inasmuch as their evaluation is based on no other information but very simple kinds of objective data, namely, either enumeration of "symmetric" cases, or observed frequencies on "analogous" trials. By contrast one should call "subjective" the probabilities evaluated in other cases, where such oversimplifying circumstances do either not exist or exist but associated with some more information to be taken into account in order to form a definite degree of belief on every single event.

3.- What "subjective" means for Subjectivists

For Subjectivists, such a distinction is illusory. Every evaluation of a probability is based on the whole of the available information, consisting on objective data, but only a subjective judgement may guide us in selection what information to consider as relevant for our purposes and how to let it influence our belief.

Even in the cases where one accept the so-called objective probabilities (e.g. ratio of white balls, or observed frequency of their occurence by drawings) it is the subjective decision to admit such information and nothing else as relevant, and to make use of it in the ordinary ways, that transforms objective data (as an objective ratio of balls, or an objective observed frequency) into a probability, which is therefore subjective, just as well as in every other case.

Is it legitimate to distinguish, notwithunderstanding, different "kinds" of probability (e.g., classical or logical, and statistival or physical, etc.)?

In a sense, eve n a Subjectivist (like I.J. Good; see 1968 Salzburg Colloquim, 4-th Discussion; Proceedings to appear) may agree.

The point is to accept that:

> "To assert that something exists is to assert that it cannot be misleading to say that it exists";

in such spirit I agree with the following:
" Now de Finetti would disagree with me because he thinks that to talk about physical probabilities is misleading. Therefore he says they do no exist. But I think it is hardly misleading to say that they exist and at least more misleading to say they do not

I agree in the sense the question is here well posed. The disagreement exists, but it does only concern the compared convenience and dangerousness of calling "right" or "wrong" a meaningless sentence about the "existence" of an imaginary entity, and not a question of principle. Why I consider misleading to speak of the "existence" of special (e.g. physical) probabilities may be seen for instance in my Salzburg paper: "Initial probabilities: a prerequisite for any valid induction", in Synthese, 19 (1969) (in press). Moreover, further information allows ordinarily to go ahead with distinctions between formerly undiscernible cases (insurers, for instance, stick to statistical data for average-guessings, but are careful in trying to individualize every single risk). And much more complex combinations of circumstances, of vague reminiscences and analogies, of more or less persuasive explanatory models, are always playing a role for the assessment of all probability beliefs. (See also n. 8).

4.- Coherence and immediacy

It has been said (n.2) that the (subjective) probability is the degree of belief as evaluated by someone making the best use of all his information and of his own skill. That creates a problem which I suppose to be particulary important from the psychological point of view and in connection with psychological experimentation.
A skilful assessment of probability requires at least an intuitive understanding of the elements of probability theory (e.g., that in a partition probabilities must add up to one), and some time and effort to consider together and to check for coherence the probabilities given to several events. At the other hand, it may also seem chiefly of interest to analyse the most genuine, immediate, uncorrected answers, to be considered as undistorted expression of the authentic, deep-rooted belief.

The emphasis on the two opposite sides may vary according to the aim of any particular research. In my opinion, however, the most instructive and fruitful modality is that one, where anybody is warned and instructed about the opportunity of paying attention in order to correct and improve his assessments through a careful comparison and revision for coherence and reasonableness; of course, without external advice or influence as for the direction of his amendments. According to this viewpoint, I think that the most suitable level of experimentation and of discussion is usually that one concerning <u>the genuinely personal beliefs of any individual, revised by himself for coherence etc. at the best of his knowledge and ability.</u> A set of probability assessments vitiated by not removed incoherences existing at a level the assessor is able to master, is not a valuable subject for research nor a significant test concerning the capabilities of this person. It may only serve to investigate some side aspects, like when the analysis of a speech is expressely made taking into account unintentional lapses; such lapses may be significant for psychoanalitical purposes but would be misleading if regarded as a part of the intentional message and used for its interpretation.

At any rate, when mistakes or strange evaluations are found, the most interesting task for psychologists is to locate their origin (as for an engineer to locate the failure giving rise to misfunctionings, or for a physician to diagnose the cause of apparent symptoms). It is interesting, e.g., to distinguish errors in understanding the problem, in its mathematical interpretation, in the manipulation of mathematical formulas (explicitly performed, or simply followed by imagination), in the final numerical computions (or guesses concerning their results).

Even more concerned with the aim of a well pondered assessment of probabilities (not only as for coherence, but also for a careful weighing of the available evidence) are methods in which different assessors are repeatedly asked to think again and possibly revise their own belief after communication of the beliefs of other persons, and maybe of the reasons given by them (e.g., "Delphi Method" , Shuford). The so arising adaptive process for individual beliefs in mutual interaction should be, I suppose, a very beautiful and farreaching task for psychologists interested in the "true" SP problem (even if at the opposite utmost as for immediacy).

5.- On side-problems

Some subjects discussed in this Conference are good exemples of side-problems (and will clarify why I call them side-problems, noway diminishing the value they may have per se). There may be various systematic causes of more or less systematic distorsions in the assessment of probabilities, owing to different circumstances.

There may be errors in measurement or guessing about objective data, to be used as a basis for a probability evaluation. For the case of the frequency, by different ways of observation, we listened to a report by Manz. Similar results should probably hold also for guesses on ball-ratios etc. (e.g., guessing the ratios of different colours of candies in a transparent vessel).

These are side-problems because they concern the objective data, irrespective of any possible application to probability assessments.

Another systematic distortion seems to arise by expressing probability evaluations in the numerical scale (Fricke; implicitly, see also in Lehmann): the probabilities whose "true" value should be x receive a "distorted" value, $y=f(x)$ (whose graph has a shape like $y = f(x) = \frac{1}{2}(1+(2x-1)^3)$) We avoid the formulation in which $x=OP$ is the "Objective probability" and $y=f(x)=SP$ is the "Subjective probability", since we dismissed such notions (or terminology) in n.2. There is no harm however to use in experiments x="ratio of white balls" and $f(x)$="betting quotient announced by the subject".

The use of any $f(x)$ (not identically $f(x)=x$) is openly untenable for anybody if only his attention is called upon absurd implications of such distortion, with reference to any interpretation of probability.

It is sufficient, for instance, to exemplify this way:

Tossing this die, you agree to consider the six faces as equally probable;

Each face has, in your belief, a probability of 20% (e.g.), as you said;

That implies, in subjectivistic terms, that
Y ou are ready to pay 6 x 2o = 12o to receive 1oo whatever happens;
Or, in statistical frequentist terms, that
You expect that, in average, points 1,2,3,4,5,6 will appear 2o times each in every sequence of 1oo trials;
Or, in classical terms, that
You believe that 20% of the 6 faces, that is 1.2o faces, are marked 1, and the like for faces marked 2,3,4,5,6.
If you dislike that, you have no way to escape but by admitting f(1/6)=1/6 (not 0.20) (and, in general, f(x)=x for every x in (0,1)

6.- Experiments on real versus artificial events

A striking contrast seems exist between the kinds of experiments preferred and performed, although with similar purposes, by psychologists at one side and by mathematicians-statisticians-economists at the other.
The latter group is inclined to place confidence on large-scale experiments related to real-life questions: meteorological forecastings (Murphy), price oscillations at Stock Exchange (Staël von Holstein), football tournaments (de Finetti, Winckler), applications to Educational Testing (Coombs, Shuford, de Finetti), not to speak about work on Marketing, Management, Operations Research, etc., only indirectly relevant. People participating to the experiments are in this case fully informed and somehow trained about the aims and methods, that are as simple and direct as possible. The rewards, if any, are carefully gradated according to decision-theoretical requirements: the recourse to the notion of utility is contemplated only when it is justified by the presence of large risks (e.g., by Grayson, studying the behavior of "wildcatters", engaged in big investments for research of oil-fields)
The Delphi Method (see n.4) has been applied to even more terrible questions (expected effects of an atomic attack under various hypotheses. As for Grayson's experience, see C.J. Grayson jr., Decisions under Uncertainty: Drilling decisions by Oil and Gas Operators, Harvard Univ., Boston 1960.

The former group seems inclined to place confidence in more conventional laboratory experiments, concerning artificial schemes of games prepared by the experimenter, and elaborated by complex and sometimes formidable statistical machinery and manipulations by computers. The rewards, even if intended to yield information about the utility curve, are kept at the level of few cigarettes or candies.

Maybe such laboratory experiments by psychologists are better performed as for aspects unknown or disregarded by statisticians, but on the whole I see several points in favor of real-life questions. with this questions anybody (the experimenter as well as the subjects) is in the same position as for possibility of information and so on.

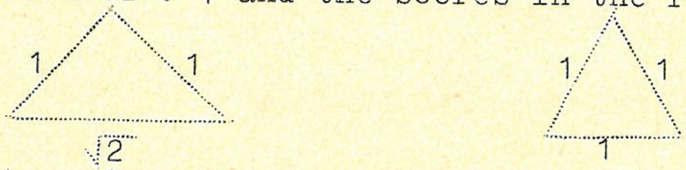
When the experience is artificial, the problem for people engaged is more or less openly to devine what the experimenter intended to test; the situation is that imagined in an amusing science-fiction novel, where an extraterrestrial animal reacted strangely to a psychological experiment because (at it was ultimately discovered) it was interested and able to detect the psychology of experimenter.

With facts like football no problem of this kind (gametheoretical situations, conscious or unconscious cheating or suspicion of cheating, etc.) can occur. The same information is available to everybody on newspapers. Similar advantages come out of the other mentioned elements of realism.

7.- Experiments on football forecastings

I will not repeat here some information, verbally given at the conference, about such experiments, hold at the University of Rome (1960-61 and 1961-62 at the Fac. of Economics; since 1966-67 at the Fac. of Sciences). As for the methods see e.g. "Does it make sense to speak of 'good probability appraisers'?, in J.Good (ed.), The Scientist speculates: An Anthology of partly-baked Ideas, Heinemann, London 1962.

As for the choice among different possible scoring rules (roughly speaking, in the sense of a "distance" between forecast and outcome), I agreed (with Stael von Holstein) there may be reasons of preference according to a feeling of more or less serious "mistakes". In the football example, it would be better in this sense to ascribe a higher value to the distance between winning and losing a match than to the distance of both such results from tying (e.g., adopting the representation on a rectangular isosceles triangle rather than the usual one, on an equilateral triangle: the distances were in the ratio $\sqrt{2}$: 1 and the scores in the ratio 2 : 1);



I did not so for the sake of simplicity (not to adopt special scales for particular applications). In another sense, indeed, the choice is indifferent (and it is in this sense I maintained this point, without contradiction with what asserted just below): in order to induce people to express genuinely their beliefs (inasmuch that is made their most convenient choice) all decisiontheoretically consistent scoring rules are equivalent.
As for analyses and conclusions of psychological character and interest, some research has been attempted, but I had no opportunity to take advantage or advice or cooperation from psychologists, as it would be most helpful, indispensable and welcome. Tentatively, I can say that a fictitious player using the average of probabilities assessed by all the real ones is always near the top positions (only few are better of as score), and the attempts to detect systematical causes of bad scores (like: overrating the supported team) were unsuccessful, so that it might be said (without claiming the existence of any underlying meaning) that things are "as if" individual beliefs were yielded by "random deviations" from a "reasonable" average belief.
Winckler's conclusions on experiments of the same nature seem similar. At any rate, there is much more to do in such kind of analysis, for this and other fields of applications. Are psychologists willing to help?

8.- Facts (objective) and their significancy (subjective)

The fundamental point for the whole thesis (or conception) maintained here is a strong distinction between what is objective, namely the observed facts, and what is not, namely the beliefs about their interpretation and significancy.
This is the essence of all what has been said till now, and some remarks to be finally added are not intended to introduce a new idea but further instances of application of the same idea.

The objective observed facts of the past do not carry, per se, any information concerning the future. It is only our subjective judgement, based on more or less reasonable ideas or conjectures about the circumstances in which these facts are produced, that makes us distinguish what ones among the observable features may be considered as significant, so that it seems reasonable to foresee that these will be preserved in the future whilst for the others no such expectation seems valid.

An experiment mentioned by Vlek seems to give a good exemple for illustration of the thesis in a wide context (and with reference to the plea for realism). Suppose, someone is steadily informed about the outcomes in a series of trials on a given phenomenon, and thus on the fluctuations of the frequency, and is asked to assess the probability he is giving to a success in the next trial.

I do not remember wheather and how the conditions were specified in Vlek's exemple. The information was graphically given by a diagram of the frequency, and I think it is rather immaterial for our discussions to know more precisely the system. To avoid excessive vagueness it might be imagined that $y=f(n)$=the frequency of successes on the last 1oo trials (that is, on those with index $n-100$, $n-99$, $n-98, \ldots, n-2, n-1$).

Whatever the size and shape of the fluctuations of the frequency could be, according to the past observations, it is only the additional knowledge about the nature of the phenomenon observed (if it is a real problem) that enables to think what features should be more or less tentatively considered as significant and what others not, and why.

In some cases it could be reasonable to disregard every information because all fluctuations are considered as accurring by chance: so it is unwise to gamble with the hope of detecting regularities in a series of trials, e.g., at the roulette, if we agree on "randomness". In other cases it might seem reasonable to expect some "seasonal" effect, and one may feel confident in a periodical return of fluctuations of roughly the same shape. Different kinds of phenomena may suggest variations of irregular lengths (e.g., if related to periods of fine or bad weather), or with a trend to decreasing (e.g., death after a surgical operation for

which tecnical improvement are in progress), or with correlation between successive trials, e.g. in a contest, when after a success someone is more confident and has more probability to repeat a good performance in the next trial. (Markov chain) And so on, and so on.
This is the reason I feel rather puzzling the situation of somebody asked to express a forecast knowing the diagram of the fluctuations but being uninformed about the real nature of the phenomena observed(or supposed to be observed).
Here, as in all cases where realistic information is lacking, I fear no reasonable attitude exists for people participating in such experiments but to try to detect what answer should better agree with what the experimenter had in his mind.
There do not exist problems, I maintain, for which realistic information is needless. Or, more precisely, they exist only in textbooks on probability. And not because of any actual reason, but only since everything is made there tautologically implicit in the bare wording.

List of Prospective Participants of Research Conferences on Subjective Probability and Related Fields

(x = participated in 1969)

J.K. Clarkson
Dept. of Psychology
University of Sheffield
Sheffield S102TN, England

J. Ekel
Dept. of Psychology
Univers. Warsaw
Krakowski Przedmiescie 24/26
Warsaw, Poland

J. Grabitz
Institut für Sozialwissensch.
der Universität Mannheim
D 68 Mannheim 1
Schloss

J.B. Kidd
Central Electricity Generating Board
Haslucks Green Road, Shirley
Solihull, Warwickshire
England

x J. Kriz
Institut für Höhere Studien
Stumpergasse 56
A-1060 Wien
Österreich

x W. Manz
Dept. of Psychology
University of Sussex
Arts Building, Falmer
Brighton, Sussex
England

x L.F.W. De Klerk
Dept. of Psychology
University of Leyden
Rijnsburgerweg 169
Leyden, Netherlands

x B. Genser
2 Hamburg 13
Von-Melle-Park 6
Psychol. Institut d. Universität
Germany

x H. Jungermann
Institut für Psychologie
Technische Hochschule
6100 Darmstadt
Neckarstrasse 6

x G. Kleiter
Psychol. Institut d. Univ. Salzburg
5010 Salzburg
Franziskanergasse 1
Österreich

B. Larsson
Dept. of Educational and Psychological Research
School of Education, Fack
S-20045 Malmö 23, Sweden

x D. Marks
Dpt. of Psychology
University of Sheffield
Sheffield S102TN, England

B. Matalon
C.R.E.D.O.C.
3o Rue d'Astorg
Paris VIII
France

x H.C. van der Meer (Miss)
Inst. voor Algemene Psychologie
Oude Boteringestraat 34
Groningen, The Netherlands

R.F. van Naerssen
Psychologisch Laboratorium
University of Amsterdam
Herengracht 510
Amsterdam, The Netherlands

F. Ölander
Economic Research Institute
Box 6501
113 83 Stockholm, Sweden

L.D. Phillips
1 Holford Road
Hampstead
London N.W. III, England

Prof. Anatol Rapoport
Technical University of Denmark
Dept. of Statistics and
Operations Research
Lyngby, Denmark

A.F. Sanders
Institute of Perception
Kampweg 5
Soesterberg, The Netherlands

x C.A.S. Staël von Holstein
Strindbergsgatan 53
S-115 31 Stockholm
Sweden

M. Sarell
Paul Baerwald School of
Social Work
The Hebrew University
Jerusalem, Israel

M. Strizenec
Inst. für Experimentelle Psychol.
SAV
Bratislava, Czechoslovakia

A. Tversky
Dept. of Psychology
The Hebrew University
Jerusalem, Israel

T. Tyszka
Dept. of Psychology
University of Warsaw
Krakowski Przedmiescie 24/26
Warsaw, Poland

x A.H.C. Van der Heijden
Psychological Institute
University of Leyden
Rijnsburgerweg 169
Leyden, The Netherlands

H.D. Schmidt
Inst. f. Psychol. d. Humbold Univ.
x-1o2 Berlin C 2
Oranienburgerstr. 18

x C.A.J. Vlek
Psychological Institute
University of Leyden
Rijnsburgerweg 169
Leyden, The Netherlands

M. Yaari
Dept. of Economics
Hebrew University
Jerusalem, Israel

Prof. D.V. Lindley
Dept. of Statistics
University College London
Gower Street
London W.C. I, England

x W.A. Wagenaar
Institute of Perception
Kampweg 5
Soesterberg, The Netherlands

Dr. V. Sarris
Psychologisches Institut
4 Düsseldorf
Himmelgeisterstr. 127

x M. Burisch
2 Hamburg 1
Rosenstrasse 9
Germany

Prof. Luciano Daboni
Dept. of Psychology
University of Trieste
Italy

x D. Wendt
Psychol. Inst. d. Univ. Hamburg
2 Hamburg 13
Von-Melle-Park 6
Germany

x Prof B. de Finetti
Istituto Matematico
Università di Roma
Via Poggio Catino 7
00199- Roma, Italy

Prof. Dr. G. Pfanzagl
c/o Inst. f. Sozialpsychol.
der Universität zu Köln
5 Köln-Lindenthal
Haedenkampstr. 2 , Germany

x G. de Zeeuw
Psychological Laborator.
University of Amsterdam
Binnengasthuis
Amsterdam, The Netherlands

x H. Imkamp
Rappstrasse 7
Hamburg 13
Germany

Mrs. Seija Hannukainen
Dept. of Psychology
University of Jyväskylä
Finland

Prof. Dario Fürst
Dept. of Psychology
University of Firenze
Italy

I. Latakos
London School of Economics
University of London
England

x R. Kakuska
2 Hamburg 13
Fontenay 6
Germany

x G. Lehmann
Wirtsch. u. Sozialwiss. Fak.
(Seminar f. Wirtsch. u. Sozialpsychologie)
Universität Göttingen
Germany

x R.E. Schaefer
Psychol. Inst. Univ. Mannheim
68 Mannheim
Schloß
Germany

x N. Mai
Psychol Inst. Univ. Mannheim
68 Mannheim
Schloß
Germany

x C. Laemmerhold
Rechenzentrum Univ. Mannheim
68 Mannheim
Schloß
Germany

x J. Werner
Sozial-Psychiatrie
Univ. Heidelberg (6900)
Luisenstr. 5
G ermany

x D. Revenstorff
Max Planck Inst. f. Psychiatrie
8 München 23
Kraepelinstr. 1o
Germany

G. Gekeler
Psychol. Institut d. T.U. Berlin
1 Berlin 33
Lassenstr. 11-15

x D. Freitag
21 Hamburg 90
Sudemannstr. 31 a
Germany

x H. Rüppell
2o91 Roydorf
An Halloh 6
Germany

R.W. Goldsmith
Max-Planck-Inst. f. Psychiatrie
8 München 23
Kraepelinstr. 1o
Germany

F. Faulbaum
2 Hamburg 73
Hüllenkamp 79

List of Publications in the Field of Subjective Probability as Submitted by the Participants

G. Kleiter: "Krise der Signifikanztests in der Psychologie (will be printed 1969 in: Jahrbuch für Psychol., Psychotherapie u. medizinische Anthropologie)

Kriz, Jürgen: "Der Likelihood-Quotient zur Erfassung des subjektiven Signifikanzniveaus", Forschungsber. No 9, Inst. f. Höhere Studien, Wien, Juli 1967

"Subjektives Signifikanzniveau im zwei-Stichprobenvergleich für Alternativmerkmale", Z. Psychol., 1968, 174, 231-244

" Über die Unabhängigkeit zwischen subjektiven und objektiven Wahrscheinlichkeiten", Forschungsbericht, No 17, Inst.f.Höhere Studien, Wien, Juli 1968

"Subjektive Wahrscheinlichkeiten und Entscheidungen" , Dissertation, Wien, Nov. 1968

B. Larsson: "Reliability and Subjective Probabilities." Didakometry, School of Education, Malmö, No. 11, 1966

"Bayes Strategies and Human Information Seeking". Lund: CWK Gleerup, 1968

"A Bayesian Marking Procedure" Didakometry and Sociometry, 1, 1969 (in press)

B. Leonardz, and Staël von Holstein: " A Comparison Between Bayesian and Classical Methods for Estimating Unknown Probabilities " (1967)

D.F. Marks: "The consistency of subjective probability". Paper read at the 1968 Annual Conference of the B ritish Psychological Society. Abstract in Bull. Brit. Psychol. Soc. Vol. 21, 115

"Further evidence that SP's are not always consistent". Paper read at the 1969 Annual Conference of the British Psychological Society.

H.C.van der Meer: "The influence of instruction in a two-choice probabilistic learning task under partial reinforcement." Acta Psychol. 196o, 17, 357-376

" Utiliteit en subjectieve waarschijnlijkheid." Rapp. Psychol. Lab. Univ. Utrecht, Br. 61-01, 1961

"Decision-making: The influence of probability preference, variance preference and expected value on strategy in gambling." Acta Psychol. 1963, 21, 231-259

"De perceptie van de waarschijnlijkheid van tussenmenselijke relaties." Ned.Ts.Psychol. 1965, 2o, 73-94

"Decision-making I. Prestatiemotivatie en kansvoorkeur onder spelen prestatieorientatie." Rapp. Psychol. Lab. Univ. Utrecht, Nr. 66-01, 1966

"Decision-making II. Enkele persoonlijkheidscorrelaten van risiconemend gedrag." Rapp.Psychol.Lab.Univ.Utrecht,Nr.66-02,1966

"Decision-making III. De relatie tussen risicobereidheid, prestatiemotivatie en tijdbeleven."
Rapp.Psychol.Lab.Univ.Utrecht,Nr.66-03,1966

"Decision-making IV. Risico en waarneming I",
Rapp.Psychol.Lab. Univ. Utrecht,Nr.66-04,1966

"Besluitvorming I. Enkele persoonlijkheidscorrelaten van risiconemend gedrag." Ned.TS.Psychol. 1966, 21, 642-660

"Besluitvorming II. De relatie tussen risicobereidheid, prestatiemotivatie en tijdbeleven."

Decision_making: Need for achievement and probability preferences under chance and skill orientation." Acta Psychol.167, 26, 353-372.

Staël von Holstein, C.-A.S.: "The Effect of Learning on the Assessment of Subjective Probability Distributions"(1968)

"The Assessment of Discrete Subjective Probability Distributions- An Experimental Study" (1969)

"Two Techniques for Assessment of Subjective Probability Distributions- An Experimental Study" (1969)

"Some Problems in the Practical Application of Bayesian Decision Theory" (1969)

Vlek, Ch.: "The use of probabilistic information in decision making" Report 009-65, Psychol. Inst., Univ. Leyden, December 1965

Vlek, C.A.J., and Beintema, K.A.: "Subjective Likelihoods in posterior probability estimation." Report No.E 014-67, Psychol.Inst.Univ.Leyden, January 1967

Vlek, C.A.J., and Van der Heijden, L.H.C.: "Subjective Likelihood functions and variations in the accuracy of probabilistic information processing." Report No. E 017-67,Psychol. Inst.Univ.Leyden,November 1967

Vlek, C.A.J., and Van der Heijden, L.H.C.: "Alternative data-generating models for probabilistic information processing."Organisational Behavior and Human Performance, 1969 (in press).

W.A. Wagenaar: "Sequential Response Bias in psychofisical experiments."Perception & Psychofysics 1968, 364-366

"Inter-dependency of successive responses in a psychofysical experiment" IZF Report No. 1968-21 (Submitted for publication)

"The recognition of randomness in visually presented binary lists."IZF Report No.1968-15 (Submitted for publication).