Justifying Bayesianism by Dynamic Decision Principles

Peter P. Wakker

Leiden University Medical Center, The Netherlands

March, 1999

ABSTRACT. As yet, no general agreement has been reached on whether the Bayesian or the frequentist (Neyman-Pearson, NP) approach to statistics is to be preferred. Whereas Bayesians adhere to coherence conditions of de Finetti, Savage, and others, frequentists do not consider these conditions normative and deliberately and knowingly violate them. Hence further arguments, bringing more clarity on the disagreements, are warranted. Providing such arguments, by refining the coherence conditions, is the purpose of this paper. It invokes recent arguments from the economic literature demonstrating that some seemingly self-evident principles for dynamic decision making have a surprising implication for static decisions: They imply Bayesianism. These principles are forgone-event independence (independence of past counterfactual events, often called consequentialism in decision theory and known as the conditionality principle in statistics), dynamic consistency (what is optimal at some given time point is independent of the time point at which that is decided), and two other conditions. Thus, a more sensitive diagnostic tool is obtained for identifying the disagreements between Bayesians and frequentists. If a frequentist does not mind violating Bayesian coherence, a Bayesian can now ask a follow-up question: Which of the dynamic principles will the frequentist give up? The debate may lead either to Bayesianism or to better implementations of non-Bayesian models in dynamic decision situations and to better non-Bayesian methods for updating information.

The diagnostic tool sheds new light on NP hypothesis testing. NP theory requires that statistical procedures are laid down before data are observed. It adheres to dynamic consistency but violates forgone-event independence. Forgone-event independence, however, is so natural that NP practitioners adhere to it and observe the data before deciding on a statistical procedure. They are thus led into violations of dynamic consistency.

KEYWORDS: Foundations of statistics; Bayesian statistics; nonexpected utility; likelihood principle; consequentialism; dynamic consistency.

1. Introduction and Definitions

Bayesianism assumes that choices between alternative courses of action under uncertainty can be decided by expected utility. That is, probability-weighted averages of consequence-utilities are maximized or, equivalently, probability-weighted averages of losses are minimized. This decision model has been extensively used in economics and underlies the Bayesian approach to statistics. A behavioral foundation was provided by Ramsey (1931), de Finetti (1937), von Neumann & Morgenstern (1944), Savage (1954), and others. They showed that Bayesianism holds if and only if some principles for rational decision making hold. They also derived Bayesian updating from such principles. To many, including the author, the principles seem rational and thus justify Bayesianism.

Contrary to some early expectations, Bayesianism has not (yet) become the dominant paradigm in statistics and economics. In statistics, the classical Neyman-Pearson approach continues to prevail. In economics, objections against the founding principles have been raised by Allais (1953), Ellsberg (1961), and others. Systematic violations have been found empirically (Kahneman & Tversky 1979) and defended normatively (Machina 1982). In Figure 2, a classical example will be presented in which the majority of people violate expected utility. These findings have led to a flurry of alternative, non-Bayesian models, called non-expected utility models, during the last 20 years (Harless & Camerer 1994). The study of these models, exploring different avenues for risk behavior, is obviously worthwhile for descriptive purposes because people in actual behavior do deviate from normative optimality. The study is also useful for normative purposes because it deepens our understanding of risk behavior. A profound discussion of Bayesianism is provided by Broome (1991, Chapter 5).

When nonexpected utility's performance in dynamic situations was studied, a surprising discovery was made. It turned out that simple, and commonly considered self-evident, conditions for dynamic choice cannot be reconciled with nonexpected utility. That is, these dynamic conditions imply Bayesianism for static choice. Thus, unexpected new support for Bayesianism was obtained. The point was first brought up by the philosopher Burks (1977, Chapter 5), a work that has unfortunately remained unknown in statistics and economics. The ideas became known in economics only after the independent discovery by Hammond (1988). I think that the arguments of Burks and Hammond deepen the earlier defenses of Bayesianism by de Finetti, Savage, and others and provide the most convincing foundation for Bayesianism presently available.

This paper reviews and unifies various discussions of the Burks/Hammond argument and shows its implications for the foundations of statistics. Section 2 presents basic definitions of decision under risk. Section 3 presents the Burks/Hammond argument, and Section 4 reviews the various reactions to this argument. If not leading to Bayesianism, the argument may lead to better implementations of non-Bayesian models in dynamic decision situations. Section 5 shows how the Burks/Hammond argument sheds new light on the Bayesian/Neyman-Pearson discussion of statistics and Birnbaum's (1962) derivation of the likelihood principle. Some subjective opinions are expressed in the concluding Section 6.

2. Decision under Risk and the "Independence" Preference Condition

The most important behavioral principle for Bayesianism is Savage's sure-thing principle. For risk, where probabilities are given, Savage's condition comes down to "independence" of von Neumann & Morgenstern (1944). As the discussion of these two conditions proceeds along the same lines, we consider only risk, for simplicity of the presentation. The arguments presented hereafter also pertain to decision under uncertainty (for example, as in Ellsberg 1961). The logical interrelations of the conditions mentioned, and the history of their inventions, are described by Fishburn & Wakker (1995). Formally, we consider lotteries, i.e. finitely-supported probability distributions over a set of *consequences* (e.g., amounts of money). $\succcurlyeq$ denotes a preference relation over lotteries that describes optimal decision making, i.e. $P \succcurlyeq Q$ means that lottery P is chosen over lottery Q.

Expected utility means that there exists a function U, called *utility function*, from the consequences to the reals such that a lottery P, yielding consequence x_j with probability p_j, j=1,…,n, is evaluated by $\sum_{j=1}^{n} p_j U(x_j)$. This summation is the *expected utility* of the lottery. Of two lotteries, the one with the higher expected utility is preferred. The function –U can be called a *loss function*. Statistics often adopts a loss function, the expectation of which is to be minimized.

Independence holds if

$$P \succcurlyeq Q \iff \lambda P + (1-\lambda)C \succcurlyeq \lambda Q + (1-\lambda)C \quad (2.1)$$

for all $0 < \lambda < 1$ and lotteries P, Q, C. Here $\lambda P + (1-\lambda)C$ is the “probabilistic mixture" of P and C that assigns λ times the probability of P plus (1-λ) times the probability of Q to each consequence. C abbreviates “common." The condition is illustrated through equivalence of choices in Figs 1a and 1e; Figure 2 provides a numerical example. Independence means that replacing an “ingredient" Q of the right-hand mixture by a better ingredient P improves the overall mix. It excludes interactions between the ingredients of the mixture.

Independence, together with some other conditions (“completeness," “transitivity," “continuity"), implies expected utility, hence Bayesianism. It is also implied by expected utility hence characterizes it. This paper concentrates on the role of independence (the sure-thing principle for uncertainty) and the other conditions are assumed throughout without further mention. Definitions and discussions of the other conditions can be found in Broome (1991).

3. A Defense of Bayesianism through Dynamic Principles

This section presents the Burks/Hammond defense of expected utility, modified in the light of the many subsequent discussions that were raised in the literature; see Figure 1. Square nodes indicate *decision nodes* where the decision maker decides where to go, circle nodes indicate *chance nodes* where chance decides where to go, and arrows in decision trees indicate preferred paths. Figure 1 uses five stages of decisions and deviates somewhat from the reasonings by Burks and Hammond. A separation into five stages, similar to Figure 1, has been considered by Cubitt, Starmer, & Sugden (1998). In all the following figures, we assume that the time between separate stages in the trees is so small that it has no physical impact other than determining the order of the nodes.

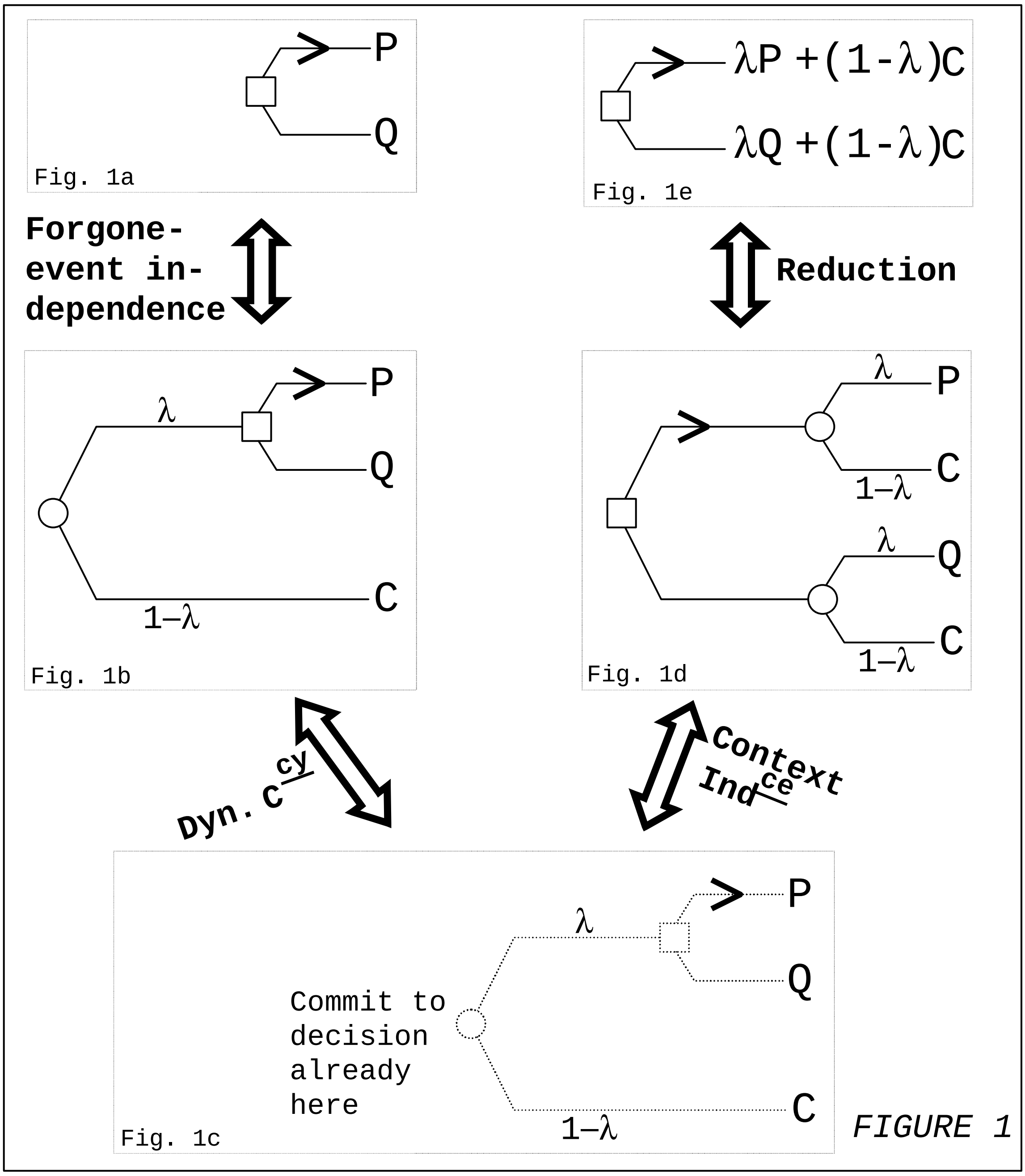

FIGURE 1

Assume the preference in Fig. 1a and consider the decision node in Fig. 1b. At that decision node, the choice options and their consequences are the same as in Fig. 1a. Hence, *forgone-event independence* requires that the choices in Figs 1a and 1b coincide. It means that at the decision node in Fig. 1b, the lower (1-λ) probability branch is irrelevant to what is best at the decision node. Indeed, if one arrives at the decision node, then the lower branch is a nonexisting event that once in the past was a possibility but is not anymore. It cannot have any physical effect on the consequences of choices at the decision node. A somewhat circular “ceteris paribus" condition underlies this assumption. That is, it is assumed that the prehistory of the prior resolution of uncertainty in Fig. 1b did not change anything “relevant" at the choice node, so that in every relevant aspect the future in Fig. 1a is the same as at the choice node in Fig. 1b. Similar ceteris paribus assumptions underlie the other dynamic principles presented later. A tongue-in-cheek formulation of forgone-event independence could be: “Don't believe in ghosts."

Fig. 1c is the same as Fig. 1b, only the decision maker must commit beforehand, before the resolution of uncertainty at the first chance node, to a decision at the future decision node. There is no possibility for later deviation from the prior commitment. *Dynamic consistency* requires that the choices in Figs 1b and 1c coincide. What is best at the decision node should not depend on the time point at which the decision maker thinks about it. A tongue-in-cheek formulation of the condition could be: “Don't be a split personality."

In Fig. 1c, the decision node is not depicted at the time point at which the decision is actually made; hence the dashed lines. A proper decision tree of the decision situation in Fig. 1c, with decision nodes at the time of decision, is presented in Fig. 1d. The upper branch depicts the prior commitment of going up in Fig. 1c, the lower branch depicts the prior commitment of going down. Figs 1c and 1d depict the same situation, and hence should be treated the same. This requirement is called *context-independence*. Violation suggests that the decision maker values lotteries not only by their intrinsic nature but also by their appearance and context. A tongue-in-cheek formulation of the condition could be: “Don't judge a book by its cover."

In Fig. 1d, the uncertainty is resolved in two stages, where first the λ or $1-\lambda$ branch is chosen and next the belonging lottery is played. A probability distribution over consequences is generated as depicted in Fig. 1e. Hence Figs 1d and 1e should be treated the same. This condition is called *reduction*. For uncertainty where no probabilities are given, the condition has been known as the collapsing of consecutive event nodes. A tongue-in-cheek formulation of reduction could be: “Don't violate the probability calculus."

The four conditions together imply equivalence of the choices in Figs 1a and 1e, i.e. independence and thus, given the other conditions assumed throughout this analysis, Bayesianism. This constitutes a modern version of the Burks/Hammond argument.

Other reasonings using decision trees have been used in the early decision analysis literature. Examples are Raiffa's (1961, his “strict dominance” entails equivalence of our Figs 1a and 1d, and his “objectively identical" claim is based on reduction) and Schlaifer (1969, Section 4.4.5, his “substitution” entails our forgone-event independence and dynamic consistency). Strategic equivalences between decision trees are also studied in game theory (compare Elmes & Reny, 1994, Figure 3.1 with the equivalence of Fig. 1b and 1d). A difficulty in the interpretation of early literature is that authors have usually discussed only one or two of the dynamic conditions explicitly, assuming the other conditions implicitly. In addition, verbal expressions of the conditions are often ambiguous. Another difficulty is that the terminologies vary widely. For example, Hammond (1986) uses the term consequentialism (currently mostly used for forgone-event independence, Machina 1989) for equivalences of Figs 1b, 1c, and 1d with 1e. Burks (1977) calls the equivalence of Figs 1a and 1c invariance (Axiom IV.a), and the equivalence of Figs 1c and 1d with 1e normal-form equivalence.

4. Explanations for Violations of the Dynamic Principles

The result of the preceding section poses a challenge to non-Bayesians. Every non-Bayesian has a question to answer, namely, which of the dynamic principles he or she will violate --it must be one at least. That holds in particular for all the readers who favor a choice down in Fig. 2a and a choice up in Fig. 2e. The example is a variation on the well-known Allais paradox (Allais 1953). The choices depicted are majority choices. In an experiment among sophisticated subjects (mostly econometricians at Tilburg), 18 of 26 subject chose up in Fig. 2a and down in Fig. 2e, 5 chose up in both figures and 3 chose down in both figures. No-one chose up in

EXAMPLE **(most critical case for independence): Test yourself.** Do you have the (nonBayesian) preferences as in Figs 2a and 2e (M = million)? Then you violate independence (λ = 0.11, C yielding $0 for sure). Determine where between Figs 2a, …, 2e your preference switches from lower to upper, i.e., which dynamic principle you violate.

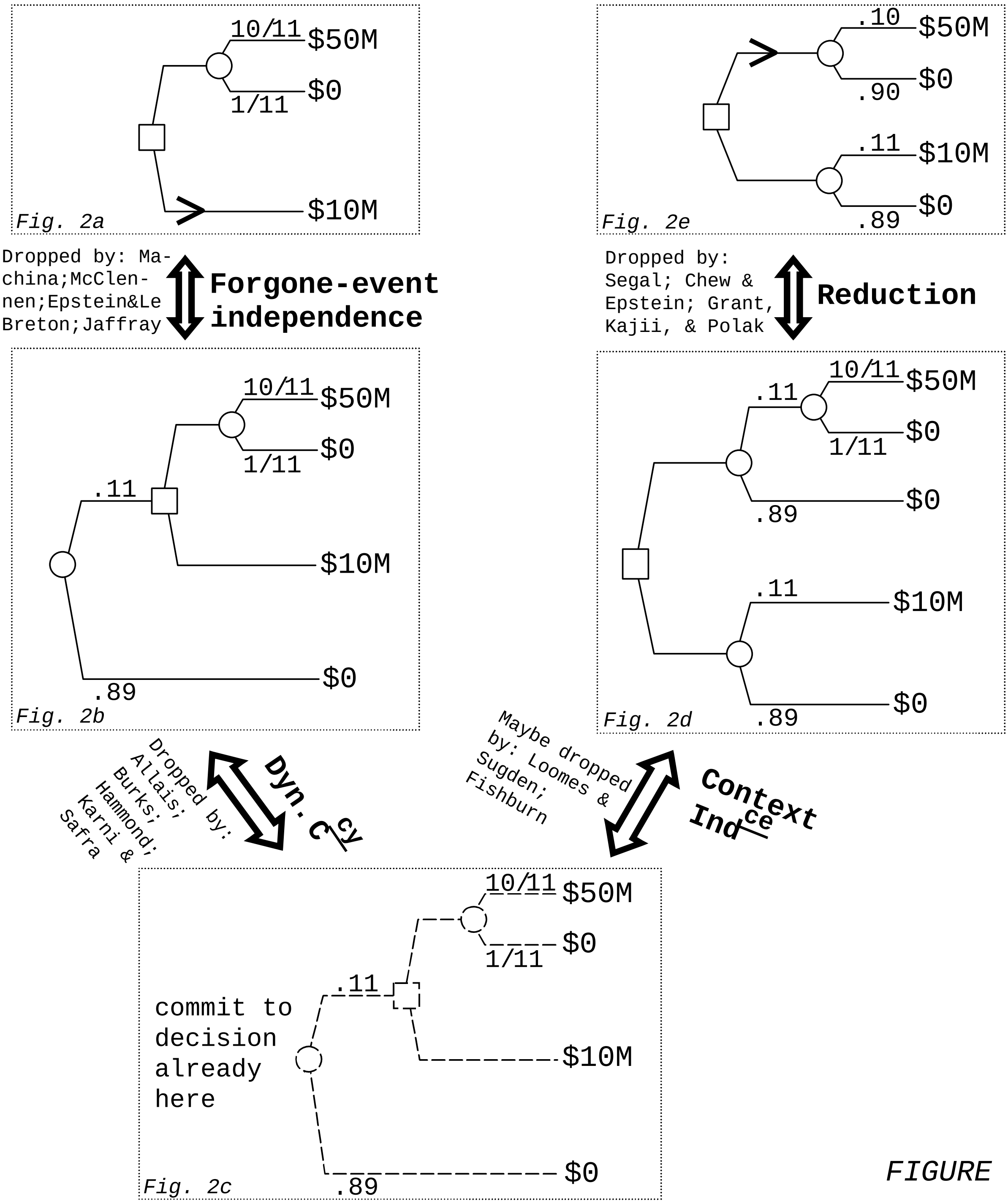

FIGURE 2

Fig. 2a and down in Fig. 2e.[1] Although the focus of this paper is normative, let me mention two empirical studies of dynamic decision principles. Cubitt, Starmer, & Sugden (1998) and Busemeyer, Weg, Barkan, & Ma (1998) tested similar dynamic principles and found that dynamic consistency was the most-violated condition. A difference between their experiments and the presentation in this paper (and belonging lectures) is that in this paper each of the figures b-e is described as a variation of the preceding figure.

GENERATION OF THE PROBABILITIES

To clarify issues regarding the nature of the probabilities, correlations, the timing of the stages and decisions etc., the random events underlying the probabilities in Figure 2 are specified. For Fig. 2a, imagine a box containing one blue card, taken from eleven blue cards numbered 1-11. It is not known which card has been taken and all cards are equally likely. The upper lottery yields \$50M (M = million) if the number on the card in the box is 10 or lower, and \$0 if the number is 11. The lower lottery always yields \$10M, independently of the card in the box. A decision maker must choose.

Fig. 2b is like Fig. 2a, but specifies a history that preceded Fig. 2a. The one card in the box was actually taken randomly from 100 cards numbered 1–100, with the first eleven cards blue and the remaining 89 cards red. If a red card had been drawn, no choice would have been offered and \$0 would have resulted. However, a blue card has been drawn, its color but not its number has been shown to the decision maker, and the card has been put into the box. Then, one minute after that, the choice as depicted in Fig. 2a is offered. This whole procedure is known to the decision maker, in particular he knows that \$0 would have resulted if the card had been red.

Fig. 2c is like Fig. 2b but with one difference, i.e. the decision maker must announce his decision a little earlier. One minute before the card is drawn, so just before the color is known, the decision maker must announce, without possibility for later change, whether at the decision node it should go up or down.

Fig. 2d is an alternative illustration of the same decision situation as in Fig. 2c. The decision node illustrates the options “announce up” and “announce down” of Fig. 2c. In particular, Fig. 2d maintains the linking of card numbers and consequences.

---

[1] The stopping-rule used in this experiment is not easily explained here.

Fig. 2e depicts a "collapsed" variation of the decision situation of Fig. 2d. When the card has been drawn immediately its number is given, so that the informations about color and number come at the same time. Other than that, the situation in Fig. 2e is the same, with the same linking of card numbers and consequences. The assumptions regarding cards imply, in particular, that the lotteries depicted in Fig. 2e are not independent. Transitivity implies that a choice between two lotteries should not be affected by whether or not they are independent. This will be discussed in further detail hereafter, in particular in Footnote 2.

With underlying events (cards) specified, we can formulate Hammond's (1988) definition of consequentialism. It requires that the only relevant aspect of a choice option is which consequences it assigns to which card numbers. The condition implies the same choices in Figs 2b, 2c, 2d, and 2e.

Let us next discuss reactions of non-Bayesians to the Burks/Hammond argument. For each principle there have been non-Bayesian authors arguing for abandoning that principle but maintaining the other ones.

ABANDONING REDUCTION

Arguments for abandoning reduction were provided by Segal (1990, his term compound independence designates same choices in Figs 1a/2a and 1d/2d). He argued that multi-stage gambles should be distinguished from single-stage gambles. Luce (1990) studied this approach for descriptive reasons (he used the term consequence monotonicity to designate same choices in Figs 1a/2a and 1d/2d). Most studies that abandon reduction consider decision trees with physically relevant time elapsing between the stages of the tree. Then the timing of the resolution of uncertainty can be relevant (Chew & Epstein 1989), for instance if there are hidden (not-modeled) decisions before the resolution of uncertainty (Kreps & Porteus 1978), or if information has value for other reasons (Grant, Kajii, & Polak 1997) . Such physically relevant timing is not assumed in our analysis.

ABANDONING CONTEXT-INDEPENDENCE

It is not easy to explain violations of context-independence in the card-example of Figure 2. Fig. 2d depicts exactly the same decision situation as Fig. 2c, is only a different way of depicting this same situation. Fig. 2d is more in agreement with the principles of decision trees. Context-independence may be violated due to "regret" effects combined with ignored correlation in Fig. 2d, as follows. If, after the upper choice in Fig. 2c, the zero consequence results (card no 11), then the decision maker may feel extra miserable thinking that he has missed a sure $10M (regret). That regret can, however, likewise be felt in Fig. 2d if upper is chosen and the card drawn is no 11. Possibly people misperceive the lotteries in Fig. 2d as if independient (which they are not), in which case a similar regret would not occur in Fig. 2d.

In general, regret considerations lead to what has sometimes been called context dependence. That is, the evaluation of consequences of a lottery is not based on their intrinsic nature but is contaminated by psychological thoughts about the different choice alternatives. Such contaminations lead to intransitivities.[2] Normative defenses of regret have been suggested by Loomes & Sugden (1982) and Fishburn (1988).

An alternative explanation for violations of context-independence occurs in descriptive settings (Kahneman & Tversky 1979, “isolation"): Fig. 2c makes the common part C (the lower 0.89 branch) of the lotteries salient, and hence subjects ignore it so as to simplify their decision. In Fig. 2d, the common part is less salient and hence subjects adopt other decision heuristics such as considering the probabilities as similar and going by the best consequence. I am not aware of normative defenses of such decision heuristics. They could be explained by bounded rationality.

If context-independence is considered self-evident for normative purposes, then the term dynamic consistency can also be used for equivalent choices in Figs 2b and 2d. This is how I presented the first lectures on this topic, with Fig. 2c only described verbally and identified with Fig. 2d. Discussions with audiences made me realize that the, supposed, difference between Figs 2c and 2d makes many people want to switch preference.

[2] Assume that a choice between two lotteries can depend on whether lotteries are independent or correlated. Then at least one preference deviates from what certainty equivalences (i.e. the amount of money for sure that is indifferent) predict and a preference cycle can be constructed. Formally, such preferences are outside the scope of this paper because we have assumed transitivity. Hence, it is formally permitted to suppress information on correlation in the figures as is done.

Abandoning Dynamic Consistency

Whereas context-independence and reduction are often violated descriptively (Camerer & Ho 1994), dynamic consistency and forgone-event independence seem to be most critical for normative purposes. Several advocates of nonexpected utility have favored abandoning dynamic consistency. An example is Allais (1953), who argued for the distinction between ex ante and ex post choice; see also Burks (1977, p. 307/308). Karni & Safra (1990) have explicitly and strongly favored abandoning dynamic consistency.[3] We call the resulting approach *sophisticated choice*. It has also been considered by Strotz (1956) in a context without uncertainty and with sequential decisions. It can imply that a doctor, one minute before opening the envelope containing a test result, declares that after a positive test result operation is best, however, one minute after having opened the envelope and having found a positive test result indeed, declares that not operating is best.

Abandoning Forgone-Event Independence

McClennen (1985) and Machina (1989) have argued for abandoning forgone-event independence (“resolute choice"). McClennen argued for an internal preference of an agent for choosing a prior plan and then following it without need of an extraneous commitment device. Machina argued that preferences are different at the decision node in Fig. 1b than in Fig. 1a because they are affected by “risks borne in the past." Also Epstein & Le Breton (1993) favor this approach. They point out (p. 11/12) that violation of dynamic consistency implies the existence of aversion to costfree information, an implication also pointed out by Wakker (1988).[4] A pragmatic advantage of resolute choice is that strategies in dynamic decision problems can be identified with single-stage probability distributions over consequences and then traditional static decision theories can still be invoked.

General Comments

Strotz (1956) considered “time inconsistencies," i.e. differences between analogs of Figs 1a and 1c. He then discussed the two possibilities of, first, precommitment (dynamic consistency) and, second, sophisticated choice. Strotz, like Hammond

---

[3] They use the term for a weaker condition (only avoiding “myopic choice”) than defined here.

[4] Forgone-event independence was described in Section 4, “first objection,” of the latter reference.

(1976, p. 162/163), considered precommitment only viable if an extraneous precommitment device is available.

Let me mention another reason why abandoning dynamic consistency seems more natural than abandoning forgone-event independence. If dynamic consistency is abandoned, then decisions are affected by events that at the moment of decision may still become reality in the future. Adapting Machina’s (1989) terminology, the risks are still borne today. It is psychologically plausible that these affect valuations. If forgone-event independence is abandoned, then decisions are affected by events that at the moment of decision are already known never to come into existence and that can therefore already be forgotten. In Machina’s terms, these are risks borne in the past.

To avoid misunderstanding, let me emphasize that forgone-event independence only requires independence of *counterfactual* events from the past. It does not impose any independence of past consumption. It is obvious that events that have happened in the past, such as seeing a movie, can and should affect your preferences today and thus also affect the movie you like to see today. Forgone-event independence allows for the latter kind of dependence on past consumption.

Apart from reduction, all conditions have in common that they exclude contamination of the evaluation of an actual consequence by counterfactual consequences. Regret concerns counterfactual acts of the decision maker (what if I had done otherwise), dynamic consistency concerns counterfactual future resolutions of uncertainty, and forgone-event independence concerns counterfactual past resolutions of uncertainty.

5. The Classical Neyman-Pearson Theory of Hypothesis Testing

Neyman-Pearson (*NP*) hypothesis testing is non-Bayesian and must therefore violate at least one of the dynamic principles presented in the preceding section. This section explains which one. Examples similar to the one used hereafter have been discussed by Cox (1958), Pratt (in Birnbaum 1962), Berger & Wolpert (1984), Bernardo & Smith (1994), and others. For another kind of “ancillarity paradox,” see Foster & George (1996).

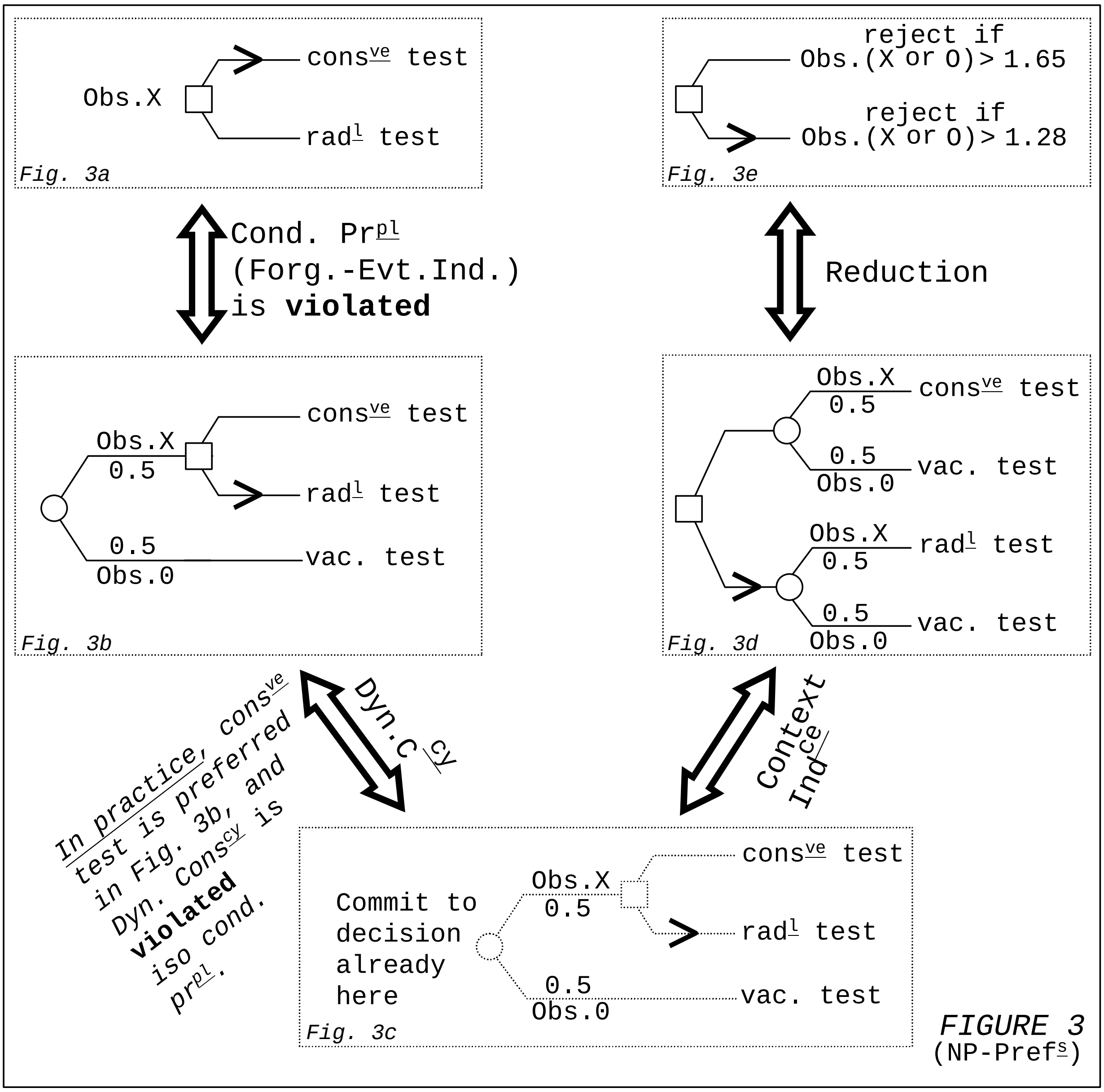

FIGURE 3 (NP-Pref^s)

The discussions in the statistics literature have often adopted a primitive concept of “evidential meaning" (for a criticism, see LeCam's discussion in Berger & Wolpert 1984). I identify that with consequences which are primitives in decision theory; they can include prescriptions for future decisions. It has sometimes been argued that statistics is an inferential task which should be distinguished from decision making (for a discussion see Schweder 1988). The decisions considered in this paper should be interpreted in a broad sense and incorporate all statistical actions such as “deciding to reject the null hypothesis" or "deciding to choose x as estimation." The consequences of decisions can accordingly be anything, such as facing a future task of predicting other parameters. In particular, consequences need not be quantified. The following discussion will not depend on any assumption regarding the nature of consequences.

Assume henceforth that $X =^d N(\mu,1)$, i.e. X is a normally distributed statistic with unknown expectation μ and known variance 1. The null hypothesis is H_0: $\mu = 0$ and the alternative hypothesis is H_1: $\mu > 0$. We will invoke, in various contexts, three tests:

Radical test: reject H_0 if $X > 1.28$ ($p = 0.10$);
Conservative test: reject H_0 if $X > 1.65$ ($p = 0.05$);
Vacuous test: never reject H_0 ($p = 0.00$).

We assume throughout that a level of significance $\alpha = 0.05$ has been generally accepted. Two different observations are discussed hereafter, *Obs.X* (observe X) and *Obs.0* (always observe 0). The second observation can occur for instance if the measurement instrument for observing X did not function. Consider Figure 3, which depicts preferences prescribed by NP theory. In Fig. 3a, the conservative test is preferred because the p-value of the radical test exceeds the significance level. The conservative test is uniformly most powerful.

Next consider Fig. 3b. Assume that there is a 0.5 probability that the measurement instrument will not function, in which case Obs.0 results. With 0.5 probability the measurement instrument will function and X is observed. Forgone-event independence implies that at the decision node in Fig. 3b, one can forget about the, by then counterfactual, lower branch of a malfunctioning measurement instrument, assume that the decision node is the beginning node of the decision tree, and therefore analyze the decision in Fig. 3b as in Fig. 3a. It is crucial that the random variable describing up or down at the first chance node is an *ancillary statistic*, i.e. its probability distribution is independent of the unknown parameter μ. Otherwise there would be a relevant difference between Fig. 3a and the decision node in Fig. 3b, being different information on μ, and forgone-event independence would not apply.

The approach just described constitutes the *conditionality principle* from statistics, i.e. one can condition on an ancillary statistic before planning the statistical procedure (Fisher 1935). Evans, Fraser, & Monette (1986), generalizing Birnbaum (1962), have demonstrated that it is mainly this condition which implies the likelihood principle. The present paper is in fact a defence of the likelihood principle. Once that principle is accepted, the Bayesian approach is natural although alternatives exist (Berger & Wolpert 1984, Chapter 5).

The classical NP theory prescribes a different procedure in Fig. 3b. It requires planning from the prior perspective, before observations have been made (Bérod 1994, Goutis & Casella 1995). That is, the analysis goes through Fig. 3c. Here the prior preference is for the radical test, which from the prior perspective now has a p-value of 0.05 and is the uniformly most powerful test. Indeed, the NP primitives, size, power, etc., only have meaning from the prior perspective. NP theory requires that in Fig. 3b the decision is followed that is optimal from the prior perspective. That is, dynamic consistency is prescribed and the radical test is chosen in Fig. 3b. It implies that the conditionality principle, thus forgone-event independence, are violated.

Context-independence and reduction have rarely been discussed in statistics. Some related arguments are found in Berger & Wolpert (1984), where criticisms of sufficiency are described that come down to rejecting reduction (Section 3.6.4 and Lane's "post-randomization" argument in the discussion). Hence these principles will not be discussed here. Let me only explain Fig. 3e, which summarizes the case. The statistic Obs.(X or 0) assigns probability 0.5 to the 0 consequence and, conditionally on a nonzero consequence, is normally distributed. The upper test rejects H_0 if the statistic Obs.(X or 0) exceeds 1.65, the lower one if Obs.(X or 0) exceeds 1.28. The lower one has size 0.05 and is uniformly most powerful, hence is preferred by NP theory. Because dynamic consistency, context-independence, and reduction are usually not discussed explicitly in the statistics literature, it cannot be determined unambiguously whether the conditionality principle is supposed to entail them or not. The restricted definition chosen in this paper, where the conditionality principle only entails forgone-event independence, serves to maximally highlight the parting of ways of the Bayesian and frequentist approaches.

In summary, NP theory abandons forgone-event independence and adheres to the other principles. As pointed out by Epstein & Le Breton (1993, p. 4) it therefore belongs to the resolute choice approach advocated by McClennen (1985) and

Machina (1989). NP practitioners commonly do not follow the prescriptions of the theory. They first observe the data ---not only ancillary statistics but in fact all the data. Only then, when all the data is available, is the final statistical procedure chosen. Often researchers, in the case of hypothesis testing, favor rejection of the null hypothesis and therefore, when several tests are available, choose that test that leads to rejection for the actually observed data. In regression, people first collect the data and then search for significant regression weights (McCloskey, 1985). Hence, people try to satisfy forgone-event independence as long as it can be suggested that the adopted decision agrees with *some* prior perspective decision. NP practitioners, unknowingly, try to satisfy forgone-event independence and to violate dynamic consistency as long as the latter violation can remain undetected.

Let me further discuss the discrepancy between the NP theory and practice. Epstein & Le Breton (1993) suggest that dynamic consistency (p. 11) and reduction (p. 12) are natural in statistics, and that forgone-event independence should be abandoned. I, however, sympathize with NP practitioners who observe data before committing themselves to a statistical decision. The intuitions of forgone-event independence and its implication, the conditionality principle, are simply too strong. Every practitioner will feel that it cannot matter for the quality of the information and of future decisions based on that, what statistical procedure the practitioner had in mind before the data were observed. That is, it cannot matter what the planned course of action was in the case of counterfactual events (nonobserved values of the statistic) that could have happened but are already known not to have. It is true that now the claimed levels of significance, powers, etc. are not formally correct because they do depend on the counterfactual courses of action, i.e. on the lower branches in the figures. I think that the resulting discrepancy is due to an unsound foundation of NP statistics and its primitives. Size and power do not provide good criterica for guiding statistical decisions. By giving up Bayesianism, NP statistics is committed to violations of logical principles for dynamic decisions, be it forgone-event independence or dynamic consistency. Thus, it leads into a realm of contradictory intuitions. This explains why, for most of this century, NP practitioners have violated its principles every day.

A historical explanation for the discrepancy between the Bayesian and the NP approach may be found in the writings of Neyman & Pearson (1933). NP discovered that in a single test between two simple hypotheses, one can equivalently express the optimal criterion in terms of a maximized likelihood ratio and in terms of a significance level plus maximal power given the significance level. They then took the, I think unfortunate, decision to take significance level and power, rather than likelihood ratio, as the criterion for guiding statistical decisions. Their choice can be explained by the unfortunate coincidence that significance levels and powers happen to have a frequentist meaning, whereas likelihood ratios do not, and by their reluctance to use prior probability for extending the likelihood ratio principle to composite hypotheses. By the power/significance criterion, the conservative test is optimal in Fig. 3a but not in Fig. 3b. By likelihood ratios, the conservative tests in Figs 3a and 3b are equivalent at all nonzero observations because they then generate the same likelihood ratios, and so are the radical tests. Likelihood ratios lead to the likelihood principle and fulfilment of the dynamic decision principles.

To conclude, I hope that, as long as Neyman-Pearson and Bayesian statistics continue to be used, its advocates and opponents will continue exchanging arguments so as to improve our understanding of statistics.

6. Conclusion

This paper has discussed three dynamic decision principles, forgone-event independence, dynamic consistency, and reduction of compound lotteries, plus some elementary rationality principles (context independence, transitivity, etc.) for decision under risk, decision under uncertainty, and statistics. The principles imply Bayesianism.

As a personal opinion, I find all dynamic principles normatively imperative. If, under appropriate ceteris paribus conditions, what is best depends on an event that is known to be counterfactual, or on whether what is best is declared one minute before or one minute after a resolution of an uncertainty, or on the way the decision situation is depicted, or on whether or not there is one minute between two future pieces of information, then I think the decision process cannot be rational. This leads me to favor the independence preference condition in decision under risk, the sure-thing principle for decision under uncertainty, and the likelihood principle for statistics.

Acknowledgment

The material presented in this paper has been developed during almost two decades. It started in August/September 1981 with a study of Chapter 5 of Burks (1977), brought to my attention by Frans A.J. Birrer, and described by Wakker (1986). The material has been further developed through the following lectures, with useful interactions with the audiences gratefully acknowledged. Fig. 2c was added on April 13, 1994, the specification of underlying cards (to trivialize the equivalence 2c-2d) on April 26, 1996.

- European Seminar in Mathematical Psychology, Madrid, Jan. 24 1992;
- Workshop Uncertainty and Decisions; University of Toulon and Du Var, France. March 19, 1992;
- Evening-course medical decision making, Medical Dept., Leiden University, April 14, 1993;
- CORE seminar, Louvain-la-Neuve, Belgium, February 14, 1994;
- Nordic Ph.D. course, Dept. of Economics, University of Copenhagen, March 25, 1994;
- Course medical decision making, Medical Dept., Leiden University, April 13, 1994;
- Seminar Dept. of Applied Mathematics, Delft University, The Netherlands, April 26, 1994;
- Game Theory day, CentER, Tilburg University, The Netherlands, April 19,1996;
- Seminar Dept. of Economics, University of Birmingham, April 26, 1996;
- Experimental Economics Course, Dept. of Economics, University of Maastricht, June 14 & June 28, 1996;
- Seminar, Economic Dept., University of Copenhagen, June 29, 1997;
- Sixth Valencia International Meeting on Bayesian Statistics (poster), Alcossebre, Spain, June 3, 1998;
- Seminar Fuqua School of Business, Duke University, Durham NC, USA, July 14, 1998;
- Seminar Dept. of Computer Science, University of Amsterdam, The Netherlands, February 12, 1999.

References

Allais, Maurice (1953), "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Américaine," *Econometrica* 21, 503–546.

Berger, James O. & Robert L. Wolpert (1984), "*The Likelihood Principle: A Review, Generalizations and Statistical Implications*." Lecture Notes, Monograph Series, Volume 6, Institute of Mathematical Statistics, Hayward, California; 2nd edition 1988.

Bernardo, Jose M. & Adrian F.M. Smith (1994), "*Bayesian Theory*." Wiley, New York.

Bérod, Annick Clerc (1994), "Conditional Behavior of Confidence Intervals," *Scandinavian Journal of Statistics* 21, 159–167.

Birnbaum, Alan (1962), "On the Foundations of Statistical Inference," (with discussions) *Journal of the American Statistical Association* 57, 269–326.

Broome, John R. (1991), "*Weighing Goods*." Basil Blackwell, Oxford.

Burks, Arthur W. (1977), "*Chance, Cause, Reason* (An Inquiry into the Nature of Scientific Evidence)." The University of Chicago Press, Chicago.

Busemeyer, Jerome R., Ethan Weg, Rachel Barkan, & Zhengping Ma (1998), "Dynamic and Consequential Consistency of Choices between Paths of Decision Trees," Department of Psychology, Purdue University, W. Lafayette, IN 47907.

Camerer, Colin F. & Teck-Hua Ho (1994), "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty* 8, 167–196.

Chew, Soo Hong & Larry G. Epstein (1989), "The Structure of Preferences and Attitudes towards the Timing of the Resolution of Uncertainty," *International Economic Review* 30, 103–117.

Cox, D.R. (1958), "Some Problems Connected with Statistical Inference," *Annals of Mathematical Statistics* 29, 357–372.

Cubitt, Robin P., Chris Starmer, & Robert Sugden (1998), "Dynamic Choice and the Common Ratio Effect: An Experimental Investigation," *The Economic Journal*, forthcoming.

de Finetti, Bruno (1937), "La Prévision: Ses Lois Logiques, ses Sources Subjectives," *Annales de l'Institut Henri Poincaré* 7, 1–68. Translated into

English by Henry E. Kyburg Jr., "Foresight: Its Logical Laws, its Subjective Sources." *In* Henry E. Kyburg Jr. & Howard E. Smokler (1964, Eds.), *Studies in Subjective Probability*, Wiley, New York, 53–118; 2nd edition 1980, Krieger, New York.

Ellsberg, Daniel (1961), "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* 75, 643–669.

Elmes, S. & P.J. Reny (1994), "On the Strategic Equivalence of Extensive Form Games," *Journal of Economic Theory* 62, 1–23.

Epstein, Larry G. & Michel Le Breton (1993), "Dynamically Consistent Beliefs Must be Bayesian," *Journal of Economic Theory* 61, 1–22.

Evans, Michael J., Donald A.S. Fraser, & George Monette (1986), "On Principles and Arguments to Likelihood" (with discussion) *Canadian Journal of Statistics* 14, 181–199.

Fishburn, Peter C. (1987), "Reconsiderations in the Foundations of Decision under Uncertainty," *Economic Journal* 97, 825–841.

Fishburn, Peter C. & Peter P. Wakker (1995), "The Invention of the Independence Condition for Preferences," *Management Science* 41, 1130–1144.

Fisher, Ronald A. (1935), "*The Design of Experiments.*" Oliver and Boyd, Edinburgh.

Foster, Dean P. & Edward I. George (1996), "A Simple Ancillarity Paradox," *Scandinavian Journal of Statistics* 23, 233–242.

Goutis, Constantinos & George Casella (1995), "Frequentist Post-Data Inference," *International Statistical Review* 63, 325–344.

Grant, Simon, Atsushi Kajii, and Ben Polak (1997), "Intrinsic Preference for Information," *Journal of Economic Theory*, forthcoming.

Hammond, Peter J. (1976), "Changing Tastes and Coherent Dynamic Choice," *Review of Economic Studies* 43, 159–173.

Hammond, Peter J. (1988), "Consequentialist Foundations for Expected Utility," *Theory and Decision* 25, 25–78.

Harless, David W. & Colin F. Camerer (1994), "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica* 62, 1251–1289.

Jaffray, Jean-Yves (1998), “Implementing Resolute Choice under Uncertainty.”

Kahneman, Daniel & Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Karni, Edi & Zvi Safra (1990), "Behaviorally Consistent Optimal Stopping Rules," *Journal of Economic Theory* 51, 391–402.

Loomes, Graham & Robert Sugden (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal* 92, 805–824.

Luce, R. Duncan (1990), "Rational versus Plausible Accounting Equivalences in Preference Judgments," *Psychological Science* 1, 225–234.

Machina, Mark J. (1982), ""Expected Utility" Analysis without the Independence Axiom," *Econometrica* 50, 277–323.

Machina, Mark J. (1989), "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature* 27, 1622–1688.

McClennen, Edward F. (1985), "Prisoner's Dilemma and Resolute Choice." *In* Richmond Campbell & Lanning Sowden (Eds.), *Paradoxes of Rationality and Cooperation*, 94–104, University of British Columbia Press, Vancouver.

McCloskey, Donald N. (1985), "The Loss function has Been Mislaid: The Rethoric of Significance Tests," *American Economic Review*, *Papers and Proceedings* 75, 201–205.

Neyman, Jerzy & E.S. Pearson (1933), "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London A* 231, 289–337.

Raiffa, Howard (1961), "Risk, Uncertainty and the Savage Axioms: Comment," *Quarterly Journal of Economics* 75, 690–694.

Ramsey, Frank P. (1931), "Truth and Probability." *In* "*The Foundations of Mathematics and other Logical Essays*," 156–198, Routledge and Kegan Paul, London. Reprinted in Henry E. Kyburg Jr. & Howard E. Smokler (1964, Eds.), *Studies in Subjective Probability*, 61–92, Wiley, New York. (2nd edition 1980, Krieger, New York.)

Savage, Leonard J. (1954), "*The Foundations of Statistics*." Wiley, New York. (Second edition 1972, Dover, New York.)

Schlaifer, Robert (1969), "*Analysis of Decisions under Uncertainty*." McGraw-Hill, New York.

Schweder, Tore (1988), "A Significance Version of the Basic Neyman-Pearson Theory for Scientific Hypothesis Testing," *Scandinavian Journal of Statistics* 15, 225–242.

Segal, Uzi (1990), "Two-Stage Lotteries without the Reduction Axiom," *Econometrica* 58, 349–377.

Strotz, Robert H. (1956), "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies* 23 (issue 3, June 1956), 165–180.

Von Neumann, John & Oskar Morgenstern (1944, 1947, 1953), "*Theory of Games and Economic Behavior*." Princeton University Press, Princeton NJ.

Wakker, Peter P. (1986), “Decision Trees, Cutting and Pasting” (Course notes). Department of Mathematics, University of Nijmegen, Nijmegen, The Netherlands.

Full address:

Peter P. Wakker

Department of Medical Decision Making

LUMC 1 K6-R (Leiden University Medical Center)

P.O. Box 9600

Leiden 2300 RC

The Netherlands

--31-71.526.45.70 (O)

--31-71.526.45.74 (S)

--31-71.526.67.50 (F)

Wakker@MDM.MedFac.LeidenUniv.nl

http://center.kub.nl/staff/Wakker