

# Utility and Means in the 1930s

Pietro Muliere and Giovanni Parmigiani

*Abstract.* This paper reviews the early axiomatic treatments of quasi-linear means developed in the late 1920s and the 1930s. These years mark the beginning of both axiomatic and subjectivist probability theory as we know them today. At the same time, Kolmogorov, de Finetti and, in a sense, Ramsey took part in a perhaps lesser known debate concerning the notions of mean and certainty equivalent. The results they developed offer interesting perspectives on computing data summaries. They also anticipate key ideas in current normative theories of rational decision making.

This paper includes an extended and self-contained introduction discussing the main concepts in an informal way. The remainder focusses primarily on two early characterizations of quasi-linear means: the Nagumo–Kolmogorov theorem and de Finetti’s extension of it. These results are then related to Ramsey’s expected utility theory, to von Neumann and Morgenstern’s and to results on weighted means.

*Key words and phrases:* Quasi-linear means, certainty equivalent, utility, subjective probability.

## 1. INTRODUCTION

Finding the best way of computing a “mean” of  $n$  quantities is a time-honored problem in statistics. In this paper we review approaches developed during the late 1920s and the 1930s. Our goal is twofold: to draw attention to some relatively little-known perspectives and results on computing data summaries and to highlight the similarities between some of the key results in normative theories of means and expected utility theory.

**Means and Decisions under Uncertainty.** Rational decision making under uncertainty is concerned with choosing among actions whose consequences cannot be completely anticipated. Typically, each action may result in one of many outcomes. The expected utility principle says that it is rational to choose among actions as if there were real-valued utilities and probabilities attached to each outcome and as if outcomes were

ranked based on their expected utility score. Probabilities are used as weights in the expectation.

More specifically, some of the fundamental results in expected utility theory (Ramsey, 1926; von Neumann and Morgenstern, 1944; Savage 1954) deal with finding sensible rationality requirements on preferences among actions that are satisfied if and only if one makes choices based on the expected utility score. For a survey of these results, including extensions and critiques, see Fishburn (1981).

The expected utility score attached to an action can be considered as a real-valued summary of the outcomes that may result from the action. In this sense it is a type of mean. From this point of view, normative theories of decision making are close to the theories that aim at dictating what is the most appropriate way of computing a mean. This similarity is a running theme throughout the paper. We use it to reinterpret some older results from the perspective of current utility theory.

**Moral Expectation.** The relationship between means and rational decision making has been recognized for a long time. The notion that mathematical expectation should guide rational choice under uncertainty was formulated and discussed as early as the seventeenth century. An important problem was finding what would now be called a certainty equivalent, that is, a fixed amount that one is willing to trade against an uncertain prospect, as when paying an insurance premium. Huygens (1657) is one of the early authors who

---

*Pietro Muliere is Professor of Statistics, Dipartimento di Economia Politica e Metodi Quantitativi, Università di Pavia, via San Felice 5, 27100 Pavia, Italy. Giovanni Parmigiani is Assistant Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708-0251. This work was completed while Pietro Muliere was Visiting Professor at the Institute of Statistics and Decision Sciences, Duke University.*

used mathematical expectation to consider the fair price of a lottery.

Daniel Bernoulli (1738) is generally credited for having been the first to critically examine Huygens' tenet. His argument is best understood in connection with the notorious St. Petersburg game. The game is as follows: The house flips a coin until the first head appears, and then pays  $2^i$  coins, where  $i$  is the number of flips. What is the fair price of this game? The series defining the expected value diverges. If expectation determines the fair price, no price is too small. Bernoulli suggested that one should not act on the basis of the expected reward, but on a different kind of expectation, which he called *moral expectation*. In particular, he found it more reasonable to assume that the value of monetary returns depends only on the percentage increase rather than the absolute increase. This led him to compute the moral value of the game as the expected logarithm of the monetary returns, which is finite.

The debate on moral expectation was important throughout the eighteenth century [for a review of the role of the St. Petersburg game in the history of probability see Jorland (1987)]. Laplace dedicated to it an entire chapter of his celebrated treatise (1812). Interestingly, Laplace emphasized that the appropriateness of the use of moral expectation relies on individual preferences for relative rather than absolute gains:

[With D. Bernoulli], the concept of moral expectation was no longer a substitute but a complement to the concept of mathematical expectation, their difference stemming from the distinction between the absolute and the relative values of goods. The former being independent of, the latter increasing with the needs and desires for these goods. (1812, pp. 189–190; translation by G. Jorland, 1987)

It seems fair to say that choosing the appropriate type of expectation of an uncertain monetary payoff was seen by Laplace as the core of what is today identified as rational decision making.

After Laplace, this fertile point of view lost prominence in the statistical literature for about a century. During this time scholars in mathematics, statistics and actuarial sciences studied the problem of means and certainty equivalents in depth and devised a range of ways to determine means. However, research often proceeded by proposing a mean, or a class of means, and then investigating its formal properties. At the beginning of the 1920s, the literature on means displayed a diverse collection of functional forms, tailored to the needs of specific contexts and disciplines.

**A Formula of Bonferroni.** The late 1920s and the 1930s mark a change in the understanding of the concept of mean and a return to comprehensive normative ideas. This trend can be noticed in several parts of the

scientific community, including mathematics (functional equations, inequalities), actuarial sciences, statistics, economics, philosophy and probability. Partly, but not solely, for expository purposes, we identify the change point with the work of Bonferroni (1924, 1927, 1928). One important aspect of Bonferroni's many contributions to the theory of means was to propose a unifying formula for the calculation of a number of different means from various application fields. In a 1924 article he writes:

The most important means used in mathematical and statistical applications consist of determining the number  $M$  that relative to the quantities  $x_1, \dots, x_n$  with weights  $P_1, \dots, P_n$ , is in the following relation with respect to a function  $\psi$ :

$$(1) \quad \psi(M) = \frac{P_1\psi(x_1) + \dots + P_n\psi(x_n)}{P_1 + \dots + P_n}.$$

I will take  $x_1, \dots, x_n$  to be distinct and the weights to be positive. (1924, p. 103; our translation: we modified the notation for consistency with similar expressions elsewhere in this paper)

Here  $\psi$  is a continuous and strictly increasing function. Various choices of  $\psi$  yield commonly used means:  $\psi(x) = x$  gives the arithmetic mean,  $\psi(x) = 1/x$ ,  $x > 0$  the harmonic mean,  $\psi(x) = x^k$  the power mean (for  $k \neq 0$  and  $x$  in some real interval  $I$  where  $\psi$  is strictly monotone),  $\psi(x) = \log x$ ,  $x > 0$  the geometric mean,  $\psi(x) = \exp\{x\}$  the exponential mean and so forth.

To illustrate the type of problem behind the development of this formalism, let us consider one of the standard motivating examples (Bonferroni 1924, 1927; de Finetti 1931a). The problem is from actuarial sciences. Consider a group of individuals of which  $P_1$  have age  $x_1$ ,  $P_2$  have age  $x_2$ ,  $\dots$ ,  $P_n$  have age  $x_n$ . From the point of view of an insurance company offering life insurance, an interesting way to determine the mean age  $M$  of the group is so that "the probability of complete survival of the group after a number  $h$  of years is the same as that of a group of  $P_1 + \dots + P_n = N$  individuals of equal age  $M$ " (Bonferroni, 1924). If the individuals share the same survival law  $S$ , and deaths are independent, such  $M$  satisfies the relationship

$$\left(\frac{S(M+h)}{S(M)}\right)^N = \left(\frac{S(x_1+h)}{S(x_1)}\right)^{P_1} \times \dots \times \left(\frac{S(x_n+h)}{S(x_n)}\right)^{P_n},$$

which is of the form (1) with  $\psi(x) = \log S(x+h) - \log S(x)$ .

Bonferroni's general expression (1) was included in the influential text of Darmais (1927) as well as in Bonferroni's own 1928 text. Various authors began working on characterizations of (1), that is, on finding a set of desirable properties of means that would be satisfied if and only if a mean of the form (1) is used.

Nagumo (1930) and Kolmogorov (1930), independently, characterized (1) (for  $P_i = 1$ ) in terms of four requirements:

1. continuity and strict monotonicity of the mean in the  $x_i$ 's;
2. reflexivity (when all the  $x_i$ 's are equal to the same value, that value is the mean);
3. symmetry, that is invariance to labeling of the  $x_i$ 's; and
4. associativity (invariance of the overall mean to the replacement of a subset of the values with their partial mean).

This characterization, together with some further developments, is presented in more detail in Section 2.

**Means with an End.** A different and, in a way, complementary approach stemmed from the problem-driven nature of Bonferroni's solution. Although anticipated by Del Vecchio (1910), this approach is usually associated with Chisini, who made it explicit in two papers (1929, 1930). Quoting Chisini (1929):

The search for a mean has the purpose of simplifying a given question, by substituting to many values a single summary value, and leaving the overall picture of the problem under consideration unchanged. [. . .] One should not be thinking about the mean of two or more values, but only about the mean of those values with reference to the evaluation of a quantity that depends on them. (p. 107, our translation)

In Chisini's own example, when computing the mean of two speeds, one can be interested in (say) the total traveling time, leading to the harmonic mean, or the total fuel consumption, leading to an expression depending, in Chisini's article, on a deterministic relationship between speed and fuel consumption.

In the actuarial example discussed earlier, an alternative way of calculating the mean could be motivated by leaving unaltered the probability of extinction of the group. This would lead to the expression

$$\begin{aligned} & \left(1 - \frac{S(M+h)}{S(M)}\right)^N \\ &= \left(1 - \frac{S(x_1+h)}{S(x_1)}\right)^{P_1} \times \dots \times \left(1 - \frac{S(x_n+h)}{S(x_n)}\right)^{P_n} \end{aligned}$$

This, incidentally, is still of the form (1) with  $\psi(x) = \log[S(x) - S(x+h)] - \log S(x)$ . Means calculated based on Chisini's principle are more general than (1), as we see later.

Chisini's proposal was formally developed and generalized by de Finetti (1931a), who later termed it *functional*. De Finetti regarded this approach as the appropriate way for a subject to determine the certainty equivalent of a distribution function. In this

framework, he reinterpreted the axioms of Nagumo and Kolmogorov as natural requirements for such choice. He also extended the characterization theorem to more general spaces of distribution functions. The functional approach is presented in Section 3.

**Means and Utility.** In current decision-theoretic terms, determining the certainty equivalent of a distribution of uncertain gains according to (1) is formally equivalent to computing an expected utility score. The function  $\psi$  plays the role of utility. De Finetti commented on this after the early developments of utility theory (de Finetti 1952, 1964). In current terms, his point of view is that the Nagumo-Kolmogorov characterization of means of the form (1) amounts to the reduction of the expected utility principle to more basic axioms about the comparison of externally given probability distributions.

It is interesting to compare this later line of thinking to the treatment of subjective probability that de Finetti was developing in the early 1930s (1931b, 1937). There, subjective probability is derived on the basis of an agent's fair betting odds for events. Determining the fair betting odds for an event means declaring a fixed price at which the agent is willing to buy or sell a ticket giving a gain of  $S$  if the event occurs and a gain of 0 otherwise. Again the fundamental notion is that of certainty equivalent. However, in the problem of means, the probability distribution is fixed, and the existence of a well-behaved  $\psi$  (that can be thought of as a utility) is derived from the axioms. In the foundation of subjective probability, the utility function for money is fixed at the outset (it is actually linear), and the fact that fair betting odds behave like probabilities is derived from the coherence requirement.

A different approach, based on the joint derivation of probability and utility from the same set of preferences, was at that time being developed by Ramsey (1926). In the fundamental paper *Truth and Probability*, written in 1926 and published posthumously in 1931, Ramsey writes:

The old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds that he will accept. This method I regard as fundamentally sound, but it suffers from being insufficiently general and from being necessarily inexact. It is inexact, partly because of the diminishing return of money, partly because the person may have a special eagerness or reluctance to bet. . . . In order therefore to construct a theory of quantities of belief which shall be both more general and more exact, I propose we take as a basis . . . the theory that we act in the way we think more likely to realize the objects of our desires. (pp. 172-173)

He then moves to define degrees of belief based on the

maximization of expected utility: in particular, he first finds an event with subjective probability of  $1/2$ , then uses this to determine a real-valued utility of outcomes, and finally uses the constructed utility function to measure subjective probability. Interestingly, the second step contains (at least formally) a characterization of a special case of Bonferroni's equation (1), as we argue in Section 4.

After Ramsey, several proposals have been put forward to reduce the expected utility principle to a more compelling set of axioms about primitive notions such as preferences among acts. Most influential is the work of von Neumann and Morgenstern (1944). The connection between the axiomatic treatment of utility and the choice of a mean was acknowledged by von Neumann and Morgenstern (1944), who wrote:

We are entitled to use the unmodified "mathematical expectation" since we are satisfied with the simplified concept of utility, . . . This excludes in particular all those more elaborate concepts of "expectation" which are really attempts at improving that naive concept of utility (e.g., D. Bernoulli's "moral expectation" in the "St. Petersburg Paradox"). (p. 83)

In spite of the different context and structural assumptions, the axioms of the Nagumo–Kolmogorov characterization and the axioms of von Neumann and Morgenstern offer striking similarities. Of special importance is the parallel between the associativity condition (substituting observations with their partial mean has no effect on the global mean) and the independence condition (mixing with the same weight an option to two other options will not change the preferences). The latter was not stated in these terms by von Neumann and Morgenstern, but has been at the center of the debate on axiomatizations since not long afterwards. See also Fishburn and Wakker (1993) on the history of the independence condition in axiomatic utility theory.

Some of the criticisms against expected utility as a principle of choice, and in particular against independence as a tenable rationality requirement can be understood within the framework of a more general version of quasi-linear means, known as means with weight functions. This notion was anticipated by Gini (1938), and we give a brief account of it in Section 5.

This closes our review. We do not survey all the relevant interpretations and applications of means, even within the specific time period considered. Also we have strongly emphasized the aspects of the theory of means that are closer to current ideas in decision theory, perhaps at the expense of some of the historical context. Our review is therefore both incomplete and biased: it is a search for some little-known roots of ideas that we now use in decision making.

## 2. THE AXIOMATIZATION OF QUASI-LINEAR MEANS

A *mean* is a function  $M: \mathbb{R}^n \rightarrow \mathbb{R}$ . In this section we introduce quasi-linear means and present the Nagumo–Kolmogorov characterization result. To set the stage we begin with some early characterization results about the arithmetic mean.

### 2.1 Early Characterizations of the Arithmetic Mean

To our knowledge, the earliest attempt to motivate the use of the arithmetic mean based on a set of desirable requirements on the function  $M$  is Schiapparelli's (1868, 1875). In particular Schiapparelli characterized the arithmetic mean by the following set of basic properties. Let  $(x_1, \dots, x_n) \in \mathbb{R}^n$ :

$$\text{Translativity: } M(x_1 + c, \dots, x_n + c) = M(x_1, \dots, x_n) + c, c \in \mathbb{R};$$

$$\text{Homogeneity: } M(cx_1, \dots, cx_n) = cM(x_1, \dots, x_n), c > 0;$$

$$\text{Symmetry: } M(x_1, \dots, x_n) = M(x_{i_1}, \dots, x_{i_n}),$$

where  $i_1, \dots, i_n$  is any permutation of  $1, \dots, n$ ;

$$\text{Continuity and differentiability: } M \text{ has everywhere one-valued and continuous partial derivatives } \partial M / \partial x_i, i = 1, \dots, n.$$

An alternative set of axioms was proposed by Schimmack (1909), who replaced the continuity and differentiability conditions with the simpler requirement that substituting all but one quantity by their partial mean does not change the final result. That is

$$M_{n+1}(x_1, \dots, x_{n+1}) = M_{n+1}(x, \dots, x, x_{n+1}),$$

$$\text{where } x = M_n(x_1, \dots, x_n).$$

Schimmack's formulation is more easily justified and has the feature that the axioms are independent (Beetle, 1915).

Scholars interested in functional equations carried out similar investigations. For example, Suto (1914) postulated symmetry and

$$\text{Reflexivity: } M(x, \dots, x) = x,$$

together with the requirement that

$$M(x_1 + y_1, \dots, x_n + y_n) - M(x_1, \dots, x_n)$$

depends on the values of  $y_1, \dots, y_n$  only.

Alternative functional equations, used along with reflexivity and symmetry, were proposed by Schweitzer (1915–1916), Huntington (1927), Narumi (1929), Teodoriu (1931), Matsumara (1933) and Nakahara (1936).

### 2.2 The Nagumo–Kolmogorov Characterization

Let us now reconsider Bonferroni's equation (1) and write it as

$$(2) \quad M(x_1, \dots, x_n) = \psi^{-1} \left( \sum_{i=1}^n p_i \psi(x_i) \right),$$

where  $\psi$  is a continuous and strictly increasing function and  $\sum_{i=1}^n p_i = 1$ . Following Aczél (1966) we term *quasi-linear* the means of the form (2). If the weights  $p_i$  are all equal, the mean is said to be *symmetric quasi-linear*.

A fundamental characterization of quasi-linear means was given, independently, by Kolmogorov (1930) and Nagumo (1930, 1931), in the symmetric case. We give the characterizing properties of  $M$  following Kolmogorov:

- K1. Continuity and strict monotonicity in all coordinates;
- K2. Symmetry;
- K3. Reflexivity;
- K4. *Associativity*:  
 $M(x_1, \dots, x_n) = M(x, \dots, x, x_{k+1}, \dots, x_n)$ ,  
 where  $x = M(x_1, \dots, x_k)$  and  $1 < k < n - 1, n > 1$ .

Associativity requires that replacing a subset of the observations with their partial mean does not change the overall mean. This condition is foreshadowed in Schimmack (1909) and Bemporad (1926).

**THEOREM 1 (Nagumo–Kolmogorov).** *Let  $I \subset \mathbb{R}$  be a closed and bounded interval and  $M: \cup_{n=1}^{\infty} I^n \rightarrow \mathbb{R}$ . Conditions K1–K4 hold if and only if there exists a function  $\psi$ , strictly monotonic and continuous, such that*

$$(3) \quad M(x_1, \dots, x_n) = \psi^{-1}\left(\frac{1}{n} \sum_{i=1}^n \psi(x_i)\right).$$

A characterization of (3) was given by Aczél (1948) replacing associativity conditions with

*Bi-symmetry:*

The value of  $M(M(x_{11}, \dots, x_{1n}), \dots, M(x_{n1}, \dots, x_{nn}))$  is unaltered whenever  $x_{ik}$  is switched with  $x_{ki}$ .

Further, relaxing symmetry, Aczél obtained a characterization of the general form (2). See also Aczél (1966) and Aczél and Dhombres (1989).

### 3. THE FUNCTIONAL APPROACH

#### 3.1 Functional Means

As discussed in the introduction, the functional approach to computing a mean is based on explicitly considering the problem that the mean must help solve. In particular, one must specify one quantity of particular relevance, like the probability of complete survival in Bonferroni’s actuarial example, or the total traveling time in Chisini’s example quoted in the introduction. This feature is expressed by a real-valued function  $f(x_1, \dots, x_n)$  on  $I^n$ . In Bonferroni’s example we would have

$$f(x_1, \dots, x_n) = \frac{S(x_1 + h)}{S(x_1)} \times \dots \times \frac{S(x_n + h)}{S(x_n)}$$

Define  $g(x) = f(x, \dots, x)$ ,  $x \in I$ ; if  $g$  is invertible in  $I$ , the mean of  $x_1, \dots, x_n$  is the solution  $m$  of the equation  $g(x) = f(x_1, \dots, x_n)$  in  $x$ , that is,

$$(4) \quad m = g^{-1} \circ f(x_1, \dots, x_n).$$

Different choices of  $f$  will lead to different means. For this reason, this approach has also been termed subjective (Martinotti, 1941) and relative (Martinotti, 1931; Boldrini, 1942).

Means derived within the functional approach are reflexive. This follows directly from the definition, but is not taken as a requirement or a defining feature. None of the other properties discussed in Section 2 is guaranteed to hold for functional means. Interestingly, a functional mean does not necessarily satisfy

$$\text{Internality: } \min_i x_i \leq M(x_1, \dots, x_n) \leq \max_i x_i,$$

which was considered an essential part of the definition of mean since Cauchy (1821). For an actuarial example of a functional mean that is not internal see Bonferroni (1937). The generality of the definition of  $f$ , and in particular the violation of internality, drew criticisms in the years that followed (Martinotti, 1931, 1939, 1941; Gini, 1938). As a reaction to the functional approach, extensive efforts were put into developing general formulae to encompass a large number of means. The most ambitious of these attempts, rejecting both the axiomatic and the functional approach, is Gini’s (1938), discussed in Section 5.

Dodd (1934) systematically reviewed the properties of means, identifying sets of independent properties. In particular he showed that internality, uniqueness, homogeneity, translativity and symmetry are independent.

#### 3.2 de Finetti’s Characterization

The subjectivist nature of the functional approach appealed to de Finetti (1931a). Later (de Finetti, 1964), he also suggested it as the appropriate way of choosing the certainty equivalent of an uncertain prospect in rational decision making. From this perspective he reconsidered the axioms of the Nagumo–Kolmogorov result, interpreting them as sensible requirements for decision in the subjectivist perspective. We discuss this further in Section 4.

In addition, de Finetti (1931a) extended the characterization from simple distributions to distributions with compact support. Consider a real-valued random variable  $X$ , with given cumulative distribution function  $\Phi$  and let  $\mathcal{F}$  be a real-valued functional on the space of distribution functions, generalizing the function  $f$  considered by Chisini. Let  $I_m(x)$  be the indicator of the set  $[m, \infty)$ . If there exists a unique real number  $m$  solving

$$(5) \quad \mathcal{F}(\Phi) = \mathcal{F}(I_m),$$

$m$  is called the “mean of  $\Phi$  for the evaluation of  $\mathfrak{F}$ .” In other words, the mean is chosen so that one is indifferent between the given cumulative distribution and a distribution concentrating all of its mass on the mean. We denote means obtained from (5) by  $M_{\mathfrak{F}}(\Phi)$ . Like the means defined in (4),  $M_{\mathfrak{F}}$  is reflexive by definition, that is,  $M_{\mathfrak{F}}(I_m) = m$  irrespective of  $\mathfrak{F}$ .

Considering Bonferroni’s actuarial example again, let  $\Phi$  be the cumulative distribution of age in the group of individuals considered, and  $N$  the total number of individuals. Then

$$\mathfrak{F}(\Phi) = \exp \left\{ N \int_I \log \frac{S(x+h)}{S(x)} d\Phi(x) \right\}.$$

To extend the Nagumo–Kolmogorov result to this more general context, de Finetti uses the two properties of monotonicity and associativity. Let  $\mathcal{P}_I$  denote the space of probability distributions with mass concentrated in some closed interval  $I \in \mathcal{R}$ .

D1: *Strict monotonicity*: Let  $\Phi_1, \Phi_2$  be in  $\mathcal{P}_I$ ; if  $\Phi_1(x) \leq \Phi_2(x)$  for every  $x$  (with strict inequality for at least one  $x$ ) then  $M_{\mathfrak{F}}(\Phi_1) > M_{\mathfrak{F}}(\Phi_2)$ .

According to this requirement, an individual will determine certainty equivalents preserving stochastic dominance. Monotonicity, together with definition (5), implies internality.

Consider a distribution  $\Phi$  that results from a convex combination of two distributions  $\Phi_1$  and  $\Phi_2$  each amounting to the fractions  $\lambda_1$  and  $\lambda_2$  ( $\lambda_1 + \lambda_2 = 1$ ). Formally  $\Phi = \lambda_1\Phi_1 + \lambda_2\Phi_2$ . The property of associativity requires that the mean of  $\Phi$  is unchanged if one of the two component distributions is replaced by another one with the same mean. Formally:

D2: *Associativity*: if  $\Phi_1$  and  $\Phi_2$  in  $\mathcal{P}_I$  are such that  $M_{\mathfrak{F}}(\Phi_1) = M_{\mathfrak{F}}(\Phi_2)$ , then for every  $\Phi_3 \in \mathcal{P}_I$  and  $\lambda \in (0, 1)$ ,  $M_{\mathfrak{F}}(\lambda\Phi_1 + (1-\lambda)\Phi_3) = M_{\mathfrak{F}}(\lambda\Phi_2 + (1-\lambda)\Phi_3)$ .

In addition to the intended descriptive interpretation, associativity can be understood as a restriction on indifference relations between uncertain prospects: Indifference between two uncertain prospects is preserved if both are mixed, in the same proportion, with a third prospect. We return to discussion of associativity in relation to de Finetti’s “dutch book” argument in subsection 4.2 and to von Neumann and Morgenstern axioms in subsection 4.3.

In de Finetti’s work, the characterization of quasi-linear means takes the following form.

**THEOREM 2** (de Finetti). *Let  $I$  be a closed and bounded interval.  $M_{\mathfrak{F}}$ , defined in (5), satisfies D1 and D2 if and only if there exists a function  $\psi$ , continuous and strictly monotone, such that for every  $\Phi \in \mathcal{P}_I$ ,*

$$(6) \quad M_{\mathfrak{F}}(\Phi) = \psi^{-1} \left( \int_I \psi(x) d\Phi(x) \right),$$

where  $\psi$  is unique up to a positive affine transformation.

De Finetti’s proof is sketchy. For a rigorous treatment see Regazzini (1987). In the case of compact support, the extension to distributions that are not simple is granted by the condition of strict monotonicity. Extensions to unbounded support are obtained by Chew (1983) by introducing the following regularity conditions. Take  $I$  to be unbounded. Let  $K_j$  be an increasing family of compact intervals such that  $\lim_{j \rightarrow \infty} K_j = I$  and let  $\Phi_K$  be the restriction of  $\Phi$  to  $K$ , that is, the conditional distribution  $\Phi(\cdot | K)$ .

*Continuity*: If  $\{\Phi_j\}_{j=1}^{\infty} \in \mathcal{P}_I$  converges in distribution to  $\Phi \in \mathcal{P}_I$ , and  $\Phi$  has a compact support, then  $M(\Phi) = \lim_{j \rightarrow \infty} M(\Phi_j)$ .

*Extension*: for all  $\Phi \in \mathcal{P}_I$ ,  
 $M(\Phi) = \lim_{j \rightarrow \infty} M(\Phi_{K_j})$ .

Alternatively, a stronger sufficient condition is continuity in distribution. See also Coletti and Regoli (1986).

Strict monotonicity is crucial for the representation to hold. Let  $S(\Phi)$  be the support of  $\Phi$ . Then  $\sup_x \{x \in S(\Phi)\}$  is associative and monotone but not strictly monotone and does not admit a representation of the form (6) (see Hardy, Littlewood and Pólya, 1934). Also, consistency and associativity do not imply monotonicity; for a counterexample see Coletti and Regoli (1986).

Both de Finetti’s and Chew’s results are based on the assumption of  $\sigma$ -additivity. We are aware of no instances in which de Finetti indicates a concern for developing a theory of means under finite additivity. This could be explained with the fact that the problem of means was originally taken from a descriptive standpoint. Seidenfeld and Schervish (1983) provide further discussion of issues related to both finite additivity and extensions from simple to non-simple distributions in the context of Savage’s theory.

### 3.3 Invariance Properties and Inequalities

An important area of investigation is the characterization of means satisfying invariance conditions. As mentioned, work in this direction began with Schiapparelli (1868). In the context of quasi-linear means, important contributions were made by Bonferroni (1924, 1926, 1927), Nagumo (1930), Jessen (1931), de Finetti (1931a) and Hardy, Littlewood and Pólya (1934).

Homogeneity, defined in terms of distribution functions, states that  $M(\Phi(cx)) = M(\Phi(x))/c$ , for any  $\Phi$  having support in  $\mathbb{R}^+$  and  $c \neq 0$ . If  $c > 0$ , the only homogeneous means of the form by (6) are obtained by setting  $\psi(x) = ax^k + \beta$ , with  $k \neq 0$ ,  $x > 0$  (root-mean square) and  $\psi(x) = a \log(x) + \beta$ ,  $x > 0$  (geometric mean). Here  $a > 0$ . Consequently, the class of root-mean power and geometric mean is characterized by reflexivity, strict monotonicity, associativity and homogeneity. As previously discussed, extensions to unbounded sup-

port can be obtained on the basis of suitable continuity conditions. An alternative characterization was given by Weerahandi and Zidek (1979).

A further condition often investigated is translative, that is,  $M(\Phi(x - c)) = M(\Phi(x)) + c$ , for  $\Phi$  arbitrary and  $c \in \mathbb{R}$ . The only quasi-linear translative means are obtained by setting  $\psi(x) = \alpha x + \beta$  (arithmetic mean) or  $\psi(x) = \alpha \exp(\gamma x) + \beta$  (exponential mean). The only quasi-linear mean satisfying both translative and homogeneity is the arithmetic mean. This was known as early as Bemporad (1926).

Cooper (1927), Jessen (1931) and de Finetti (1931a) compared quasi-linear means obtained using different functionals on the same probability distribution. Let  $M_{\mathfrak{F}_1}(\Phi)$  and  $M_{\mathfrak{F}_2}(\Phi)$  be quasi-linear means with different functions  $\psi_1$  and  $\psi_2$ . If  $\psi_1$  is increasing (decreasing), then  $M_{\mathfrak{F}_1}(\Phi) \geq M_{\mathfrak{F}_2}(\Phi)$  holds for every  $\Phi$  if and only if  $\psi_1 \circ \psi_2^{-1}$  is convex (concave). As an example, take  $\psi_1(x) = x^{m_1}$  and  $\psi_2(x) = x^{m_2}$ , with  $x > 0$  and  $m_1 > m_2$ . Then  $\psi_1 \circ \psi_2^{-1} = x^{m_1/m_2}$  is convex (concave) if  $m_1$  is positive (negative).

#### 4. MEANS, UTILITY AND SUBJECTIVE PROBABILITY

In the 1920s and 1930s means had an important role in the controversy over the foundation of probability. In his very influential treatise, Keynes (1921) refuted “the doctrine that the mathematical expectations of alternative courses of actions are the proper measures of our degrees of preferences” (p. 344). Borel (1924), Ramsey (1926) and de Finetti (1931b, 1937) rejoined, albeit in the context of different proposals, that overt preferences are the only sound way to measure a person’s degree of belief. Both de Finetti and Ramsey endeavored to construct theories of probability based on this principle. We briefly review the relations of these theories to the theory of means.

##### 4.1 Means and Subjective Probability in de Finetti

Consider a decision maker determining the certainty equivalent of a given probability distribution based on a functional  $\mathfrak{F}$  of particular interest for the problem. Suppose also that the decision maker considers it natural to use an  $\mathfrak{F}$  satisfying D1 and D2. On this, de Finetti and Emanuelli (1967) wrote:

In the context of comparing preferences, associativity (D2) constitutes basically a coherence condition: operating an indifferent change on a situation, its preference or indifference relation with others cannot be changed. This is why we can conclude, in the case of decisions under uncertainty, that the certainty equivalent to an uncertain situation must be expressed by an associative mean, and therefore (from the Nagumo–Kolmogorov Theorem) there must exist a suitable in-

creasing function—the utility, for which, to an uncertain situation there corresponds the prevision (i.e., the mathematical expectation) of the utilities of the possible situations. (p. 74, our translation)

We note that the word coherence is plausibly used by de Finetti in the ordinary language sense, with no direct reference to coherence of probability assignments.

In de Finetti’s view, the characterization of associative means amounts to the reduction of the expected utility principle to more basic axioms about ranking distribution functions. In the early 1930s, de Finetti (1931b, 1937) developed an approach to subjective probability. As is well known, for de Finetti, probability is measured by the notion of fair betting odds as the certainty equivalent of a bet on the outcome of the event. In this sense, the development of the notion of probability presents similarities with the functional approach to means; both are based on a primitive notion of indifference between certain and uncertain outcomes, and on a set of rationality requirements. In Theorem 2, the function  $\psi$  (the utility) is derived with probability given extraneously. Conversely, in the definition of subjective probability, probabilities are derived from basic preferences for a fixed utility function.

However, de Finetti kept separate the derivation of probability and utility. In later writings (de Finetti 1952, 1964, 1970; de Finetti and Emanuelli, 1967) he discussed explicitly the option of deriving both utilities and probability from a single set of preferences and seemed to consider it the most appropriate way to proceed in decision problems, but maintained that the separation is preferable in general, giving two reasons.

First, the notion of probability, purified from the factors that affect utility, belongs to a logical level that I would call “superior.” Second, constructing the calculus of probability in its entirety requires vast developments concerning probability alone. (de Finetti, 1952, p. 698, our translation)

##### 4.2 Ramsey’s Approach as a Representation Theorem

In his 1926 paper, Ramsey laid the foundations of subjective probability in a different way. He proposed a scheme encompassing both what will be the functional approach to means and the Dutch Book argument. In this section, we review Ramsey’s theory with the intent of highlighting the similarities with the theory of means. A discussion in the context of expected utility theory is in Fishburn (1981). Consider an uncountable set of outcomes or consequences (*worlds* in Ramsey’s terminology), designated by  $\alpha, \beta$  and so on. Outcomes are not monetary nor necessarily numerical, but each outcome is assumed to carry a value. Outcomes in the

same equivalence class are indicated by  $\alpha \sim \beta$ . The subject is assumed to have a weak order on outcome values. Strict preference is indicated by  $\succ$ .

The general strategy of Ramsey is: first, find a neutral proposition with subjective probability of 1/2, then use this to determine a real-valued utility of outcomes and finally, use the constructed utility function to measure subjective probability.

The subject is assumed to be able to make choices between options of the form: " $\alpha$  if  $E$  is true,  $\beta$  if  $E$  is not true." We indicate such option by  $O_E(\alpha, \beta)$ , dropping  $E$  when unambiguous. The outcome  $\alpha$  and the option  $O(\alpha, \alpha)$  belong to the same equivalence class, a property only implicitly assumed by Ramsey, but very important in this context as it represents the equivalent of reflexivity.

In Ramsey's definition, an ethically neutral proposition is one whose truth or falsity is not "an object of desire to the subject." More precisely, a proposition  $E$  is ethically neutral if two possible worlds differing only by the truth of  $E$  are equally desirable. The existence of one such proposition is postulated as an axiom (R1). The notion of ethical neutrality can be read as Ramsey's attempt to address what is today known as the problem of state-dependent utilities. Schervish, Seidenfeld and Kadane (1990) discuss difficulties with axiomatic approaches based on preferences (including Ramsey's) in providing a satisfactory treatment of conditions like ethical neutrality.

Next Ramsey defines an ethically neutral proposition with probability 1/2.  $P(E) = 1/2$  if for every pair of outcomes  $(\alpha, \beta)$ ,  $O(\alpha, \beta) \sim O(\beta, \alpha)$ . That preferences among outcomes do not depend on which ethically neutral proposition with probability 1/2 is chosen is built into  $O$  (axiom R2). There is no loss of generality in taking  $P(E) = 1/2$ : the same construction could have been performed with a proposition of arbitrary probability, as long as such probability could be measured based solely on preferences.

In analysing Ramsey's construction as a representation theorem, a fundamental assumption, given as axiom R6, is that for every pair of outcomes  $(\alpha, \beta)$ , there exists a unique outcome  $\mu$  such that  $O(\alpha, \beta) \sim O(\mu, \mu)$ . As  $O(\mu, \mu) \sim \mu$ , there is a unique certainty equivalent to  $O(\alpha, \beta)$ . We indicate this by  $\mu = M\{O(\alpha, \beta)\}$ .

We now list the remaining axioms:

- R2a: If  $O(\alpha, \delta) \sim O(\beta, \gamma)$  then  $\alpha \succ \beta$  if and only if  $\gamma \succ \delta$ , and  $\alpha \sim \beta$  if and only if  $\gamma \sim \delta$ .
- R3: The indifference relation between options is transitive.
- R4: If  $O(\alpha, \delta) \sim O(\beta, \gamma)$  and  $O(\gamma, \zeta) \sim O(\delta, \eta)$ , then  $O(\alpha, \zeta) \sim O(\beta, \eta)$ .
- R5: For every  $\alpha, \beta, \gamma$ , there exists a unique outcome  $\nu$  such that  $O(\alpha, \beta) \sim O(\nu, \gamma)$
- R7: Continuity.
- R8: Axiom of Archimedes.

Ramsey provides little explanation regarding the last two axioms. Their role is to make the space of outcomes rich enough to be one-to-one with the reals. Sahlin (1990) suggests that continuity should be the analogue of the standard completeness axiom of real numbers. Then it would read like "every bounded set of outcomes has a least upper bound." Here the ordering is given by preferences.

One may also conjecture that a way of formalizing the Archimedean axiom in this framework could be the following. For every  $\alpha \succ \beta \succ \gamma$ , there exist  $\mu$  and  $\nu$  such that  $O(\nu, \beta) < \gamma$  and  $O(\mu, \beta) \succ \alpha$ . Debreu (1959) and Pfanzagl (1959, 1967, 1971) offer a precisely stated set of axioms to obtain results similar to Ramsey's.

Axioms R1-R8 give the existence of a real-valued, one-to-one, utility function on outcomes, designated by  $\psi$ , such that

$$(7) \quad O(\alpha, \delta) \sim O(\beta, \gamma) \Leftrightarrow \psi(\alpha) - \psi(\beta) = \psi(\gamma) - \psi(\delta).$$

Ramsey never gave a proof. Modified versions of this result are discussed by Suppes and Winet (1955), Davidson and Suppes (1956), Suppes (1956), Debreu (1959) and Pfanzagl (1959, 1967, 1971).

According to (7), the individual's preferences are represented by a continuous (by virtue of axioms R2a and R6) and strictly monotone utility function  $\psi$ , determined up to a positive affine transformation. In particular, consistent with the principle of expected utility,

$$O(\alpha, \delta) \sim O(\beta, \gamma) \Leftrightarrow \frac{\psi(\alpha) + \psi(\delta)}{2} = \frac{\psi(\beta) + \psi(\gamma)}{2}.$$

Now, let  $\mu = M\{O(\alpha, \delta)\}$ . From axiom R6 and (7) we can write:  $[\psi(\alpha) + \psi(\delta)]/2 = \psi(\mu)$ , so that

$$M\{O(\alpha, \delta)\} = \psi^{-1}\left[\frac{\psi(\alpha) + \psi(\delta)}{2}\right].$$

Thus Ramsey's theorem characterizes the quasi-linear form.

We now briefly discuss the relationships between the axioms of the Nagumo-Kolmogorov characterization and the axioms of Ramsey. Ramsey's context is not confined to real-valued outcomes, but, for the purpose of the comparison, we take real-valued outcomes. Then  $M\{O(\alpha, \delta)\}$  is also real valued. We have argued that reflexivity is built on the assumption that  $\alpha$  must be equivalent to " $\alpha$  if  $E$  is true,  $\alpha$  if  $E$  is not true." Strict monotonicity follows from R2a by a simple proof by contradiction. Symmetry follows from axiom R2 by taking the two ethically neutral propositions  $E$  and  $\bar{E}$ . Finally, let  $\mu = M\{O(\gamma, \delta)\}$ . Then  $O(\gamma, \delta) \sim O(\mu, \mu)$ . But then, if  $E$  and  $F$  are ethically neutral,

$$(8) \quad \begin{aligned} &M\{O_E[O_F(\alpha, \beta), O_F(\gamma, \delta)]\} \\ &= M\{O_E[O_F(\alpha, \beta), O_F(\mu, \mu)]\}. \end{aligned}$$

Here we introduce iterated mixing in a way never explic-



itly suggested by Ramsey. The point of our argument is that as  $\alpha, \beta, \gamma$  and  $\delta$  are arbitrary, (8) amounts to associativity (with the proviso that  $n$  and  $k$  must be even integers in this context). Let us now consider the inverse implications. As we confine attention to real-valued outcomes, axioms R3, R7 and R8 are automatically satisfied. Axiom R1 is satisfied, as the probabilities are determined at the outset by the dimension of the function  $M$  and by the symmetry property. Axiom R6 follows from existence of the mean and reflexivity. Axiom R2 follows from symmetry and R2a from monotonicity and reflexivity. Axiom R4 can be deduced from symmetry and associativity. Finally, axiom R5 follows from continuity and strict monotonicity.

The result of (7) must be viewed in the wider context of the determination of subjective probability, which is the aim of the representation. Here, however, we discuss Ramsey's construction only insofar as it constitutes a characterization result for the use of means of the quasi-linear form.

### 4.3 von Neumann and Morgenstern's Approach as a Representation Theorem

After Ramsey, a most influential treatment of the expected utility principle was proposed by von Neumann and Morgenstern (1944). From the point of view of this discussion, their axiomatic treatment of utility can also be seen as a representation theorem in which probability is extraneous. In particular, individuals express preferences over options—denoted by  $O$ . Options consist, as in Ramsey, of uncertain outcomes, and have a discrete and finite support. Probabilities are determined externally to the problem and do not have, unlike in Ramsey, a subjective meaning. To bring out the parallel between the theory of von Neumann and Morgenstern and that of de Finetti, we report the axioms in the condensed form of Jensen (1967).

VM1. *Ordering*: The agent holds a weak ordering  $\leq$  of options;

VM2. *Independence*: If  $O_1 \leq O_2$ , then for every  $O_3$  and  $\lambda \in (0, 1)$ ,

$$\lambda O_1 + (1 - \lambda)O_3 \leq \lambda O_2 + (1 - \lambda)O_3$$

VM3. *Archimedean*: If  $O_1 < O_2 < O_3$ , there exists  $\lambda_1$  and  $\lambda_2$  in  $(0, 1)$  such that

$$\lambda_1 O_1 + (1 - \lambda_1)O_3 < O_2 < \lambda_2 O_1 + (1 - \lambda_2)O_3.$$

An agent obeying these axioms is ranking options according to the expectation of a real-valued utility function  $\psi$  on rewards, defined uniquely by the preferences, up to a positive affine transformation. For a discussion of the meaning of the axioms, see Fishburn (1981).

As pointed out by de Finetti (1952) (see also Daboni,

1984; Fishburn and Wakker, 1993), there are important similarities between this representation and the characterization results of Sections 2 and 3. In particular, the independence axiom parallels associativity. We emphasize two differences between the two settings: first, the options of von Neumann and Morgenstern are defined on completely arbitrary rewards; second, the existence of a certainty equivalent for every lottery is not postulated, as in Ramsey's or Nagumo and Kolmogorov's formulations.

Other axiomatizations of expected utility can also be reinterpreted in this light. An enumeration is beyond the scope of this paper (see Fishburn, 1970, 1982). We just note that Savage (1954) developed an axiomatization of the expected utility principle in which, as in Ramsey's, both probability and utility are based on a single preference relation. Unlike Ramsey, however, Savage determines probability first and then builds on the work of von Neumann and Morgenstern to obtain utility.

## 5. MEANS WITH WEIGHT FUNCTION

Quasi-linear means as a model of certainty equivalence have encountered criticism (e.g., Allais, 1953), and alternative paradigms have been proposed to account for some of the more prevalent violations. Generalizations of the notion of quasi-linear means can be traced to contributions of the 1930s and continue to play a very important role in this endeavour. In particular, means with a weight function, foreshadowed by Gini (1938), were later reconsidered from both an axiomatic and a functional perspective. In this section, we introduce means with a weight function and give a brief summary of some of the results most directly connected with the characterization problem.

### 5.1 Gini Means

Differing from both the functional and the axiomatic approach, Gini (1938) proposed an elaborated family of means attempting to provide a comprehensive treatment of the commonly applied procedures. An important feature of his system is the presence of a weight function depending on the observation. For simplicity, we present here only the important special case of power means with power weight function, given by

$$(9) \quad M(x_1, \dots, x_n) = \left( \frac{\sum_{i=1}^n x_i^{s+r}}{\sum_{i=1}^n x_i^s} \right)^{1/r}.$$

This family includes, for example, the power means ( $s = 0$ ) and the contraharmonic mean ( $r = s = 1$ ). Means given by (9) satisfy internality, reflexivity, homogeneity and symmetry (Farnsworth and Orr, 1986), but not associativity. Moreover they are not necessarily monotone in  $r$  (Beckenbach, 1950). Martinotti (1939)

pointed out that (9), as well as the more general family given by Gini, can be interpreted as functional means.

## 5.2 Functional Equations and Characterizations of Means with Weight Function

Axiomatic treatments of means with a weight function have been given in the literature on functional equations. In particular, Bajraktarevic (1958) defined the quasi-linear mean with weight function to be a function of  $M: \cup_{n=1}^{\infty} I^n \rightarrow R$  such that there exists a continuous strictly monotonic function  $\psi: I \rightarrow R$  and a positive valued weight function  $p: I \rightarrow R$  for which

$$(10) \quad M(x_1, \dots, x_n) = \psi^{-1} \left( \frac{\sum_{i=1}^n p(x_i) \psi(x_i)}{\sum_{i=1}^n p(x_i)} \right)$$

for all  $n$ . Here  $I$  is any open set. When  $p(x_i) = 1$  for all  $i$ , we obtain (3). A characterization of (10) was given by Páles (1986, 1987). The characterizing properties are:

- (i) Reflexivity.
- (ii) Symmetry.
- (iii) For any  $x < u < v < y$  in  $I$  there are  $n, m$  such that  $u < M(x, \dots, x, y, \dots, y) < v$ , where  $x$  appears  $n$  times and  $y$   $m$  times.
- (iv) For  $x_1, \dots, x_n, y_1, \dots, y_m$ , each in  $I$ ,

$$\lim_{k \rightarrow \infty} M(x_1, \dots, x_1, \dots, x_n, \dots, x_n, y_1, \dots, y_m) = M(x_1, \dots, x_n),$$

where each  $x_i$  appears  $k$  times.

- (v) If  $M(x_1, \dots, x_n, u_1, \dots, u_k) \leq M(x_1, \dots, x_n, v_1, \dots, v_l)$  and  $M(y_1, \dots, y_m, u_1, \dots, u_k) \leq M(y_1, \dots, y_m, v_1, \dots, v_l)$ , then

$$M(x_1, \dots, x_n, u_1, \dots, u_k, y_1, \dots, y_m, u_1, \dots, u_k) \leq M(x_1, \dots, x_n, v_1, \dots, v_l, y_1, \dots, y_m, v_1, \dots, v_l).$$

In the axiomatic treatment of information measures, several important functions, like Shannon entropy and Rényi entropy can be expressed as special cases of (10). (See Aczél and Daróczy, 1975.) Also, some measures of inequality can be derived from (10) (Bürg and Gehrig, 1978).

We conclude the section with a result of Losonczi (1973), who interprets means with weight functions as functional means. In particular, the equation

$$\sum_{i=1}^n p(x_i) V(x_i, y) = 0,$$

where the right-hand side is continuous and strictly increasing in  $y$ , has a unique internal solution, also called *implicit mean*. If in addition,  $V(x, y) = \psi(x) - \psi(y)$ , then the implicit mean has the form (10). A more general formulation is also given by Losonczi (1973). A related concept is that of deviation means, defined

by Daróczy (1972a). See also Daróczy (1972b), Losonczi (1973), Daróczy and Páles (1980, 1982) and Páles (1982).

## 5.3 Means with a Weight Function in Utility Theory

Chew (1983) proposes a generalization of Theorem 2 of de Finetti. He considered the following characterizing properties. Let  $\mathcal{P}_I$  denote the space of probability distributions with mass concentrated in some interval  $I \in R$ .

- C1. Reflexivity.
- C2. *Betweenness*: for all  $\Phi_1$  and  $\Phi_2$  in  $\mathcal{P}_I$ , if  $M(\Phi_1) < M(\Phi_2)$ , then for all  $\lambda_1$  in  $(0, 1)$ ,  $M(\lambda_1 \Phi_1 + (1 - \lambda_1) \Phi_2) \in (M(\Phi_1), M(\Phi_2))$ .
- C3. *Substitution-independence*: suppose there are  $\Phi_1, \Phi_2$  and  $\Phi_3$  in  $\mathcal{P}_I$  and  $\lambda_1, \lambda_2$  in  $(0, 1)$  such that  $M(\Phi_1) = M(\Phi_2) \neq M(\Phi_3)$  and  $M(\lambda_1 \Phi_1 + (1 - \lambda_1) \Phi_3) = M(\lambda_2 \Phi_2 + (1 - \lambda_2) \Phi_3)$ . Then, for every  $\Phi_4$  in  $\mathcal{P}_I$ ,  $M(\lambda_1 \Phi_1 + (1 - \lambda_1) \Phi_4) = M(\lambda_2 \Phi_2 + (1 - \lambda_2) \Phi_4)$ .
- C4. Continuity, as defined in Section 3.
- C5. Extension, as defined in Section 3.

**THEOREM (Chew).** *Suppose there exists an  $M: \mathcal{P}_I \rightarrow R$ . Then  $M$  satisfies C1-C5 if and only if there are continuous functions  $\psi$  (strictly monotone) and  $\alpha$  (non-vanishing except possibly at one endpoint, in which case  $\alpha \psi \neq 0$ ) on  $I$  such that for all  $\Phi$  in  $\mathcal{P}_I$ ,*

$$(11) \quad M(\Phi) = \psi^{-1} \left( \frac{\int_I \alpha(x) \psi(x) d\Phi(x)}{\int_I \alpha(x) d\Phi(x)} \right).$$

Means defined by (11) include quasi-linearity as a special case, but are not necessarily monotone or associative. Monotonicity can be achieved by the further restriction that  $\alpha(x)(\psi(x) - \psi(s))$  is monotone in  $x$  for every  $s$ . This fact was used by Chew to provide an explanation of the Allais Paradox (see Allais, 1953), without violating either transitivity of preferences or consistency with stochastic dominance or betweenness.

The relationship between associativity and the axioms of Chew is clarified by the fact that C2 and C3 imply the following: if  $M(\Phi_1) = M(\Phi_2)$ , then for every  $\lambda_1$ , there is a  $\lambda_2$  such that for every  $\Phi_3$ ,  $M(\lambda_1 \Phi_1 + (1 - \lambda_1) \Phi_3) = M(\lambda_2 \Phi_2 + (1 - \lambda_2) \Phi_3)$ . Associativity requires the further restriction that  $\lambda_1 = \lambda_2$ .

An alternative axiomatic characterization relaxing both the ordering and the quasi-linearity assumptions was proposed by Fishburn (1986). Consider a function  $v(x, y)$  describing the intensity of the difference in preference between  $x$  and  $y$ . Then  $v$  represents a generalization of  $\psi(x) - \psi(y)$  in Ramsey's (7). Natural requirements imposed by Fishburn are that  $v$  is skew-symmetric [i.e.,  $v(y, x) = -v(x, y)$ ], strictly increasing in the first argument and ratio-continuous. The mean

value of a probability distribution  $\Phi$  is defined as the unique solution of the equation

$$\int v(x, y) d\Phi(x) = 0.$$

The quasi-linear form is obtained as a special case when  $v(x, y) = \psi(x) - \psi(y)$  and the weighted quasi-linear form is obtained when  $v(x, y) = (\psi(x) - \psi(y)) \cdot \alpha(x)\alpha(y)$ . Details of the axiomatization are given in Fishburn (1986, 1988).

## 6. CONCLUSIONS

In statistics means are ubiquitous. It was not our plan to review all interesting interpretations and applications of means, even within the time period considered. We thus close by suggesting some further references. For applications of means to statistical inference, the interested reader can consult Norris (1976), who devotes special attention to moments and maximum likelihood estimation. The notion of certainty equivalent as representative income is relevant in the measurement of income inequality and social welfare functions. A concise exposition is Chew (1987), who also points to connections between rank-dependent quasi-linear means and  $L$ -estimators and between implicit weighted quasi-linear means and robust estimators. For an extensive bibliography on inequalities among means and relations with inequalities on expectations of random variables see Bullen, Mitrinovic and Vasic (1988).

## ACKNOWLEDGMENTS

The reviewer's suggestions were very valuable and led to substantial improvement. Paolo Bertoletti, Michele Cifarelli, Jay Kadane, Eugenio Regazzini, Fabrizio Ruggeri, Marco Scarsini and Teddy Seidenfeld gave useful comments on earlier drafts.

## REFERENCES

- ACZÉL, J. (1948). On mean values. *Bull. Amer. Math. Soc.* 54 392-400.
- ACZÉL, J. (1966). *Lectures on Functional Equations and Their Applications*. Academic, New York.
- ACZÉL, J. and DARÓCZY, Z. (1975). *On Measures of Information and their Characterizations*. Academic, New York.
- ACZÉL, J. and DHOMBRES, J. (1989). *Functional Equations in Several Variables*. Cambridge Univ. Press.
- ALLAIS, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica* 21 503-546.
- BAJRAKTAREVIC, M. (1958). Sur une équation fonctionnelle aux valeurs moyennes. *Glasnik. Mat.-Fiz. Astr.* 13 243-248.
- BECKENBACH, E. F. (1950). A class of mean value functions. *Amer. Math. Monthly* 57 1-6.
- BEEBLE, R. D. (1915). On the complete independence of Schim-mack's postulates for the Arithmetic Mean. *Math. Ann.* 76 444-446.
- BEMPORAD, G. (1926). Sul principio della media aritmetica. *Atti Accad. Naz. Lincei* 6 87-91.
- BERNOULLI, D. (1738). Specimen Theoriae Novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5 175-192. [English translation as Exposition of a new theory on the measurement of risk. (1954). *Econometrica* 22 23-35.]
- BOLDRINI, M. (1942). *Statistica: Teoria e Metodi*. Giuffrè, Milano.
- BONFERRONI, C. E. (1924). La media esponenziale in matematica finanziaria. *Annuario del Regio Istituto Superiore di Scienze Economiche e Commerciali di Bari AA 23-24* 1-14.
- BONFERRONI, C. E. (1926). Sulle medie di potenze. *Giornale degli Economisti e Rivista di Statistica* 66 292-295.
- BONFERRONI, C. E. (1927). Sulle medie dedotte da funzioni concave. *Giornale di Matematica Finanziaria* 9 13-24.
- BONFERRONI, C. E. (1928). *Elementi di Statistica Generale*. Litografia Gili, Torino.
- BONFERRONI, C. E. (1937). A proposito di espressioni generali per le medie. *Giornale degli Economisti e Rivista di Statistica* 77 345-356.
- BOREL, E. (1924). A propos d'un Traité de Probabilité. *Revue Philosophique* 98 321-326. [English translation as: Apropos of a treatise on probability. In *Studies in Subjective Probability* (E. Kyburg and H. E. Smokler, eds.) Wiley, New York, 1964.]
- BULLEN, P. S., MITRINOVIC, D. S. and VASIC, P. M. (1988). *Means and Their Inequalities*. Riedel, Dordrecht.
- BÜRG, R. and GEHRIG, W. (1978). Indices of income inequality and societal income. An axiomatic approach. In *Theory and Applications of Economic Indices* (W. Eichhorn et al., eds.) 309-356. Physica, Vienna.
- CAUCHY, A. L. (1821). *Cours d'Analyse de l'Ecole Royale Polytechnique, I ere partie: Analyse Algebrique*, Paris.
- CHEW, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of the income inequality and decision theory resolving the Allais Paradox. *Econometrica* 51 1065-1092.
- CHEW, S. H. (1987). Mean value. In *The New Palgrave: Dictionary of Economics* (J. Eatwell, M. Milgate, P. Newman, eds.) 422-426. MacMillan, London.
- CHISINI, O. (1929). Sul concetto di media. *Periodico di Matematiche* 9 106-116.
- CHISINI, O. (1930). Numeros indice et medias. *Schola et Vita* 5 181-190.
- COLETTI, G. and REGOLI, G. (1986). Non monotone associative means. *Metron* 44 237-245.
- COOPER, R. (1927). Notes on certain inequalities I and II. *J. London Math. Soc.* 2 17-21, 159-163.
- DABONI, L. (1984). On the axiomatic treatment of the utility theory. *Metroeconomica* 36 203-209.
- DARMOIS, G. (1927). *Statistique Mathématique*. Doin, Paris.
- DARÓCZY, Z. (1972a). Über eine Klasse der mittelwerten. *Publ. Math. Debrecen* 19 211-217.
- DARÓCZY, Z. (1972b). A general inequality for means. *Aequationes Math.* 7 16-21.
- DARÓCZY, Z. and PÁLES, ZS. (1980). Multiplicative mean values and entropies. *Colloq. Math. Soc. János Bolyai, Functions, Series, and Operator.* 35 343-359.
- DARÓCZY, Z. and PÁLES, ZS. (1982). On comparison of mean values. *Publ. Math. Debrecen* 29 107-115.
- DAVIDSON, D. and SUPPES, P. (1956). A finitistic axiomatization of subjective probability and utility. *Econometrica* 24 264-275.
- DE FINETTI, B. (1931a). Sul concetto di media. *Giornale dell'Istituto Italiano degli Attuari* 2 369-396.
- DE FINETTI, B. (1931b). Sul significato soggettivo della probabilità. *Fund. Math.* 17 298-329.
- DE FINETTI, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* 7 1-68. [English translation]

- tion as Foresight: Its logical laws, its subjective sources. In *Studies in Subjective Probability* (H. Kyburg and H. Smokler, eds.) 93-158.]
- DE FINETTI, B. (1952). Sulla Preferibilità. *Giornale degli Economisti e Annali de Economia* 11 685-709.
- DE FINETTI, B. (1964). Teoria delle decisioni. In *Lezioni di Metodologia Statistica per Ricercatori* 6 88-161. Università degli Studi di Roma.
- DE FINETTI, B. (1970). *Teoria delle Probabilità*. Einaudi, Torino. [Translated as *Theory of Probability*. Wiley, New York, 1974.]
- DE FINETTI, B. and EMANUELLI, E. (1967). *Economia delle Assicurazioni*. UTET, Torino.
- DEBREU, G. (1959). Topological method in cardinal utility theory. In *Mathematical Methods in the Social Sciences* (K. J. Arrow, S. Karlin and P. Suppes, eds.) 16-26. Stanford Univ. Press.
- DEL VECCHIO, G.S. (1910). Di un teorema sulla applicazione delle medie a serie statistiche di quantita reciproche. *Giornale degli Economisti e Rivista de Statistica* 40 559-585.
- DODD, E. L. (1934). The complete independence of certain properties of means. *Ann. of Math.* 35 740-746.
- FARNSWORTH, D. and ORR, R. (1986). Gini means. *Amer. Math. Monthly* 93 603-607.
- FISHBURN, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.
- FISHBURN, P. C. (1981). Subjective expected utility: A review of normative theories. *Theory and Decision* 13 139-199.
- FISHBURN, P. C. (1982). *The Foundations of Expected Utility*. Reidel, Dordrecht.
- FISHBURN, P. C. (1986). Implicit mean value and certainty equivalence. *Econometrica* 54 1197-1205.
- FISHBURN, P. C. (1988). *Nonlinear Preference and Utility Theory*. Johns Hopkins Univ. Press.
- FISHBURN, P. C. and WAKKER, P. (1993). The invention of the independence condition. Technical Report, AT&T Bell Laboratories, Murray Hill, NJ.
- GINI, C. (1938). Di una formula comprensiva delle medie. *Metron* 13 3-22.
- HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934). *Inequalities*. Cambridge Univ. Press.
- HUNTINGTON, E. V. (1927). Sets of independent postulates for the arithmetic mean, the geometric mean, the harmonic mean and the root-mean-square. *Trans. Amer. Math. Soc.* 29 1-22.
- HUYGENS, C. (1657). *De Ratiociniis in Ludo Aleae*. In *Exercitationum Mathematicarum*. Elsevir, Leiden.
- JENSEN, N. E. (1967). An introduction to Bernoullian utility theory. I. Utility functions. *Swedish Journal of Economics* 69 163-183.
- JESSEN, B. (1931). Bemaerkninger om konvekse funktioner og uligheder imellem middlevaredier I, II. *Matematisk Tidsskrift* B 17-28, 84-95.
- JORLAND, G. (1987). The Saint Petersburg Paradox 1713-1937. In *The Probabilistic Revolution* (L. Kruger, L. J. Daston and M. Heidelberger, eds.) 1 157-190. MIT Press.
- KEYNES, J. M. (1921). *A Treatise on Probability*. [Reprint: McMillan, London, 1973.]
- KOLMOGOROV, A. N. (1930). Sur la notion de la moyenne. *Atti Accad. Naz. Lincei* 9 388-391.
- LAPLACE, P.-S. (1812). *Théorie Analytique des Probabilités*. Paris.
- LOSONCZI, L. (1973). General inequalities for nonsymmetric means. *Aequationes Math.* 9 221-235.
- MARTINOTTI, P. (1931). Le medie relative. *Giornale degli Economisti e Rivista de Statistica* 46 291-303.
- MARTINOTTI, P. (1939). Estensioni del concetto di media. *Giornale degli Economisti e Annali di Economia* 1 624-632.
- MARTINOTTI, P. (1941). Di alcune recenti medie. *Acta della Pontificia Accademia delle Scienze* 5 113-121.
- MATSUMARA, S. (1933). Über die axiomatik von mittelbildungen. *Tōhoku Math. J.* 36 260-262.
- NAGUMO, M. (1930). Über eine klasse der mittelwerte. *Japan. J. Math.* 7 72-79.
- NAGUMO, M. (1931). On mean values. *Tokyo Buturigakko-Zasshi* 40 520-527.
- NAKAHARA, I. (1936). Axioms for weighted means. *Tōhoku Math. J.* 39 424-434.
- NARUMI, S. (1929). Note on the law of the arithmetical mean. *Tōhoku Math. J.* 30 19-21.
- NORRIS, N. (1976). General means and statistical theory. *Amer. Statist.* 30 8-13.
- PÁLES, ZS. (1982). Characterization of quasideviation means. *Acta Math. Acad. Sci. Hungar.* 40 243-260.
- PÁLES, ZS. (1986). On the generalization of quasiarithmetic means with weight functions. *Aequationes Math.* 31 324-325.
- PÁLES, ZS. (1987). On the characterization of quasiarithmetic means with weight functions. *Aequationes Math.* 32 171-194.
- PFANZAGL, J. (1959). A general theory of measurement: Application to utility. *Naval Res. Logist.* 6 283-294.
- PFANZAGL, J. (1967). Subjective probability derived from the Morgenstern-von Neumann utility concept. In *Essays in Mathematical Economics* (M. Shubick, ed.) 237-251. Princeton Univ. Press.
- PFANZAGL, J. (1971). *Theory of Measurement*. Physica, Würzburg.
- RAMSEY, F. (1926). Truth and probability. In *The Foundations of Mathematics*. 156-211. Routledge & Kegan Paul, London.
- REGAZZINI, E. (1987). *Argomenti di Statistica. Le medie*. Quaderni del Dipartimento di Matematica "F. Enriques," Milano.
- SAHLIN, N. E. (1990). *The Philosophy of F. P. Ramsey*. Cambridge Univ. Press.
- SAVAGE, L. (1954). *The Foundations of Statistics*. Wiley, New York.
- SCHERVISH, M. J., SEIDENFELD, T. and KADANE, J. B. (1990). State dependent utilities. *J. Amer. Statist. Assoc.* 85 840-847.
- SCHIAPARELLI, G. (1868). Sul principio dell'uso della media aritmetica nel calcolo dei risultati delle osservazioni. *Rendiconti del Regio Istituto Lombardo di Scienze e Lettere* 1 771-778.
- SCHIAPARELLI, G. (1875). Sur le principe de la moyenne arithmetique. *Astronom. Nachr.* 87 55-58.
- SCHIMMACK, R. (1909). Der satz vom aritmetischen mittel in axiomatischer bergündung. *Math. Ann.* 68 125-132.
- SCHWEITZER, A. R. (1915-1916). A new functional characterization of the arithmetic mean. *Bull. Amer. Math. Soc.* 22 171.
- SEIDENFELD, T. and SCHERVISH, M. J. (1983). A conflict between finite additivity and avoiding dutch book. *Philos. Sci.* 50 398-412.
- SUPPES, P. (1956). The role of subjective probability and utility in decision making. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 5 61-73. Univ. California Press, Berkeley.
- SUPPES, P. and WINET, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Sci.* 1 259-270.
- SUTO, O. (1914). Law of the arithmetical mean. *Tōhoku Math. J.* 6 79-81.
- TEODORU, L. (1931). Sur la définition axiomatique de la moyenne. *Mathematicae* 5 27-31.
- VON NEUMANN, J. and MORGENSTERN, O. (1944). *Theory of Games and Economic Behavior*. Wiley, New York.
- WEERAHANDI, S. and ZIDEK, J. V. (1979). A characterization of the general mean. *Canad. J. Statist.* 7 83-90.