Monatsh. Math. 135, 305–319 (2002)

Monatshefte für
Mathematik

Functional Characterizations of Basic Properties of Utility Representations

By

Che Tat Ng[1], **R. Duncan Luce**[2], and **János Aczél**[1]

[1]University of Waterloo, Ont., Canada
[2]University of California, Irvine, U.S.A.

Received 10 August 2001; in revised form 6 November 2001

Dedicated to Edmund Hlawka on the occasion of his 85th birthday

Abstract. Consider uncertain alternatives for which an event has two consequences (binary gambles, "gambles" for short) and over them an operation of joint receipt which need not be closed and may be non-commutative. The two structures are linked by a distributivity property called segregation and a preference order. Utility functions order nonnegative numbers to consequences and gambles. Utility representations describe how the utility of a gamble depends on the utilities of consequences and on the "weight" of the event (a number in [0,1], depending on the event). Functional characterizations give necessary and sufficient conditions, often in form of functional equations, for certain properties of representations. We first give a functional characterization of the often postulated event commutativity stating that two events can be interchanged in special composite gambles where one outcome is a consequence but the other is itself a gamble. A utility representation is separable if it is multiplicative for gambles with one consequence having 0 utility. We give three more specific characterizations of separable representations by segregation, by homogeneity and event commutativity, and by homogeneity and segregation, and show that in the last case event commutativity follows.

2000 Mathematics Subject Classification: 39B22, 91B16
Key words: Utility, gambles, non-commutative joint receipt, event commutativity, separability, segregation, homogeneity, functional equations, representation and uniqueness theorems

1. Origins of the Problem

Consider the following problem that arises in the theory of utility representations. It is stated briefly so as to get rapidly to the problems at hand, which will be reduced to some functional equations. (For a more leisurely presentation of the utility problem, see [4].) Let X denote a set of valued consequences, "goods" or "bads" (gains or losses), that have no uncertainty associated with them. Let typical elements be $x, y \in X$. Let $\mathscr{E}_{\mathbf{E}}$ be an algebra of events arising from an "experiment" or "chance phenomenon" $\mathbf{E}$ with the universal event denoted by Ω and a typical event being denoted by $C \in \mathscr{E}_{\mathbf{E}}$. Let $(x, C; y)$ denote a binary gamble, or for short in this paper "gamble," in which the holder of the gamble receives x if C occurs when

The research of Aczél and Ng has been supported in part by the Natural Sciences and Engineering Research Council of Canada grants OGP 0002972 and OGP 0008212, respectively, and that of Luce has been supported in part by the National Science Foundation grant SBR-9808057 to the University of California, Irvine.

E is executed and y if C fails to occur. If f, g denote independently realized gambles, then $(f, C; g)$ denotes a compound gamble, compound in the sense that its outcomes are, themselves, gambles. The term "independently realized gambles" has an informal meaning similar to the concept of independent experimental trials, as for a random sample, in applied statistics.

Also, for two independently realized gambles f, g, we let $f \oplus g$ denote having or receiving the pair. We speak of $\oplus$ as *joint receipt*. We treat $f \oplus g$ as a gamble and, in particular (cf. below), $x \oplus y$ as a consequence in X for $x, y \in X$.

Define $\mathscr{D}_i$ inductively as follows:

$$\mathscr{D}_0 = X,$$
$$\mathscr{D}_i = \{(f, C; g), f \oplus g : f, g \in \mathscr{D}_{i-1}, C \in \mathscr{E}_{\mathbf{E}}\} \cup \mathscr{D}_{i-1} \quad (i = 1, 2, \ldots).$$

For the purposes of this paper, we take as our domain $\mathscr{D} = \mathscr{D}_3$. This, of course, means that $\mathscr{D}$ need not be closed either under the formation of binary gambles or joint receipts.

Assume that there exists a transitive and connected preference order $\succsim$ (i.e., *weak order*) over $\mathscr{D}$. Define $f \sim g \Leftrightarrow (f \succsim g \;\&\; g \succsim f)$, $f \succ g \Leftrightarrow (f \succsim g \;\&\; \text{not}\, f \sim g)$, and $f \prec g \Leftrightarrow g \succ f$.

It is convenient to assume that the set X is sufficiently rich so that for each $f \in \mathscr{D}$, there exists $x(f) \in X$ such that $x(f) \sim f$. It is called the *certainty equivalent of* f.

We assume that there is a distinguished element $e \in X$ that is interpreted as *no change from the status quo*. Elements of $\mathscr{D}^+ := \{f | f \in \mathscr{D} \text{ and } f \succsim e\}$ are called *gains* and those of $\mathscr{D}^- := \{f | f \in \mathscr{D} \text{ and } f \precsim e\}$ are called *losses*. Note that $\mathscr{D}^+$ includes consequences (elements x of X with $x \succsim e$); gambles such as f and g, $f \succsim e, g \succsim e$; and compound gambles such as $(f, C; g) \succsim e$.

1.1. Assumptions about Gambles. Gambles are said to exhibit

- *Left-consequence monotonicity* if

$$f \succsim f' \Leftrightarrow (f, C; g) \succsim (f', C; g) \quad (f, f', g \in \mathscr{D}_1, C \in \mathscr{E}_{\mathbf{E}}, C \neq \emptyset), \tag{1}$$

- *Complementarity* if

$$(f, C; g) \sim (g, \overline{C}; f) \quad (f, g \in \mathscr{D}_1, C \in \mathscr{E}_{\mathbf{E}}), \tag{2}$$

where $\overline{C} = \Omega \backslash C \in \mathscr{E}_{\mathbf{E}}$,

- *Idempotence* if

$$(f, C; f) \sim f \quad (f \in \mathscr{D}_1, C \in \mathscr{E}_{\mathbf{E}}), \tag{3}$$

- *Certainty* if

$$(f, \Omega; g) \sim f \quad (f, g \in \mathscr{D}_1). \tag{4}$$

In what follows we assume these four properties.

Note that from complementarity and certainty, we also have the property *nullity*

$$(f, \emptyset; g) \sim g. \tag{5}$$

1.2. Assumptions about Joint Receipt. We assume that $\langle \mathscr{D}^+, \oplus, \succsim, e\rangle$ satisfies the following properties. Preference order $\succsim$, being a weak order over $\mathscr{D}$, is of course also a weak order over $\mathscr{D}^+$, and relative to that order $\oplus$ is *strictly left-monotonic increasing* in the sense that, for all $f, f', g \in \mathscr{D}_2^+$,

$$f \succsim f' \Longleftrightarrow f \oplus g \succsim f' \oplus g. \tag{6}$$

We assume that e is a *left identity*, i.e., for all $f \in \mathscr{D}_2^+$,

$$f \sim e \oplus f. \tag{7}$$

Because we do not assume the commutativity of $\oplus$, it does not follow that e is also a right identity, $f \sim f \oplus e$. The following is trivial.

Proposition 1. *Suppose* $\succsim$ *is a weak order*, $\oplus$ *is strictly left-monotonic increasing*, (6), *and* e *is a left identity*, (7). *Then left positivity holds*, *i.e.*, *for* $f, g \in \mathscr{D}_2^+$

$$f \oplus g \succsim g.$$

Note that right positivity, $f \oplus g \succsim f$, does not automatically hold.

Note also that we have stated matters in terms of left identity, left-consequence monotonicity, left monotonicity and left positivity of $\oplus$. A dual theory holds for right identity, right-consequence monotonicity, right monotonicity and right positivity of $\oplus$.

1.3. Segregation and RDU with Commutative Joint Receipts. Unless otherwise stated, from now on consequences and gambles are assumed to be in $\mathscr{D}^+$. The results for $\mathscr{D}^-$ are parallel. Those for the mixed case are more complex [4].

In [6] Luce and Fishburn posed, and answered under the following assumptions, the question: What is the form of $U(f \oplus g)$? They assumed that the structure $\langle \mathscr{D}^+, \oplus, \succsim, e\rangle$ satisfies: $\succsim$ is a weak order, $\oplus$ is commutative and monotonic in which case e is a (two-sided) identity of $\oplus$. They also assumed the following distributivity-type property, called *segregation*, that links the gambling structure over gains to joint receipts: For all $f, g \in \mathscr{D}_1^+$ and $C \in \mathscr{E}_{\mathbf{E}}$,

$$(f, C; e) \oplus g \sim (f \oplus g, C; g). \tag{8}$$

They assumed further that the preference structure can be represented by a binary rank-dependent utility (RDU) representation of the following form: there exist a function $U : \mathscr{D}^+ \xrightarrow{\text{onto}} [0, k[, (k \in]0, \infty])$, which is called a *utility function*, and a function $W : \mathscr{E}_{\mathbf{E}} \xrightarrow{\text{onto}} [0, 1]$, which is called a *weighting function*, such that

$$U(f, C; g) = U(f)W(C) + U(g)[1 - W(C)] \quad (f, g \in \mathscr{D}_2^+, f \succsim g), \tag{9}$$

$$f \succsim g \Leftrightarrow U(f) \geqslant U(g) \quad (f, g \in \mathscr{D}^+), \tag{10}$$

$$(f, C; g) \succsim (f, D; g) \Leftrightarrow W(C) \geqslant W(D) \quad (f, g \in \mathscr{D}_2^+, f \succ g), \tag{11}$$

where we have simplified the notation by defining $U(f, C; g) := U[(f, C; g)]$. Moreover, if $f, g \in \mathscr{D}_2^+$, then $U(f) \geqslant 0, U(g) \geqslant 0$, and so by (9) $U(f, C; g) \geqslant 0$, i.e., $(f, C; g) \in \mathscr{D}^+$.

Note that (11) follows immediately from (9) and (10).

For $f \precsim g$, the form of $U(f,C;g)$ follows from (9) and the assumption of complementarity, (2).

We refer to (9) as an RDU representation, to (10) as U being order preserving, and to (11) as W being order preserving.

1.4. Generalizations. This paper concerns generalizations of two of those assumptions: the commutativity of $\oplus$ and the RDU form.

Commutativity is somewhat problematic in the utility interpretation and extremely so in a psychophysical reinterpretation of the primitives [5]. In the latter case, the function U goes under the name of "psychophysical function" or "subjective intensity." For example, let (x,y) denote presenting tone intensity x to the left ear and intensity y to the right ear and $\succsim$ the judgment of comparative loudness. This information can be used to define a mathematical operation on the left component by solving $(x,y) \sim (z,0)$ for z and writing $z = x \oplus_l y$. Commutativity of $\oplus_l$ corresponds to the behavioral symmetry $(x,y) \sim (y,x)$. Empirically, this does not hold. So a non-commutative theory is needed.

In providing one, we will invoke the several left assumptions above and use segregation as stated. In the right dual theory, segregation must be stated with g on the left.

We also drop the assumption of RDU, (9), and admit a much weaker form of the utility representation that exhibits some of the consequences of RDU including event commutativity, homogeneity, and separability, which are each defined below. When there are both gambles and joint receipts, we continue to suppose that they are linked by segregation (8). Specifically, sec. 2.3 characterizes event commutativity and 3.1 does so with the added assumption that the representation is homogeneous; sec. 2.4 characterizes segregation assuming separability and 3.2 does so under homogeneity and shows that these two assumptions imply event commutativity; and sec. 3.3 shows how the characterizations in 3.1 and 3.2 relate.

In another paper we intend to reinstate the RDU representation and use some of the present results to explore its implications in the context of non-commutative joint receipt.

2. General Utility Representations

Suppose that $U : \mathscr{D}^+ \xrightarrow{\text{onto}} [0,k[, (k \in]0,\infty])$ preserves the weak ordering $\succsim$ of $\mathscr{D}^+$, (10), satisfies left-consequence monotonicity (1), and $U(e) = 0$. Note that because we have assumed the existence of certainty equivalents, $U : \mathscr{D}_i^+ \xrightarrow{\text{onto}} [0,k[$, $i = 0,1,2,3$.

Let now $W : \mathscr{E}_{\mathbf{E}} \xrightarrow{\text{onto}} [0,1]$ satisfy (11). The pair (U,W) is said to form a *utility representation of* $\langle \mathscr{D}^+, \succsim, e\rangle$ with utility function U if for $f,g \in \mathscr{D}_2^+$, $f \succsim g$, $U(f,C;g)$ is a function just of $U(f)$, $U(g)$, and $W(C)$, that is,

$$U(f,C;g) = M_{W(C)}[U(f),U(g)] \quad (f \succsim g). \tag{12}$$

This is clearly a generalization of RDU, (9). Setting $w = W(C)$, $p = U(f), q = U(g)$, $(p,q \in [0,k[)$, we can write the definition of M as

$$M_w(p,q) := U(U^{-1}(p), W^{-1}(w); U^{-1}(q)) \quad (p \geqslant q). \tag{13}$$

We will also quote (12) as "(U, W) has the representation M" and (13) as "M is defined in terms of a utility representation (U, W)".

The results in the next proposition trivially follow.

Proposition 2. *Suppose that* (U, W) *has the representation given by* (12). *Then idempotence* (3) *is equivalent to*

$$M_w(p, p) = p \quad (w \in [0, 1],\ p \in [0, k[). \tag{14}$$

Certainty (4) *is equivalent to*

$$M_1(p, q) = p \tag{15}$$

and nullity (5) *is equivalent to*

$$M_0(p, q) = q \tag{16}$$

for $k > p \geqslant q \geqslant 0$.

When dealing with $M_w(p, q)$ with all three variables free, we speak of the function M, and when dealing with it with w held fixed we speak of the function M_w. The function M is, by left-consequence monotonicity (1) and by (10), strictly increasing in the p argument for $w \neq 0$ and, by (10) and (11), it is strictly increasing in w for $p > q$. As a consequence of strict increasing in w for $p > q$, and of (16) and (15), the function M_w is an *intern* map (mean value):

$$q = M_0(p, q) < M_w(p, q) < M_1(p, q) = p \qquad (w \in]0, 1[,\ 0 \leqslant q < p < k). \tag{17}$$

The function M is also supposed to be continuous in p and w. To summarize, we assume the following about M:

Definition 1. The function M is said to be *in the class IIIC*, written $M \in \mathit{IIIC}$, if it is strictly increasing in p for $w \neq 0$, strictly increasing in w for $p > q$, intern (17), and continuous in p for each (w, q) and in w for each (p, q) $(0 \leqslant w \leqslant 1,\ 0 \leqslant q \leqslant p < k)$.

The continuity and monotonicity in p and in w yields the *joint continuity* of M in (w, p) for each q.

We do not assume regularity conditions on q for now, but sometimes, depending on the particular applications, it is natural to impose them.

2.1. Event Commutativity. We say that a structure $\langle \mathscr{D}^+, \succsim, e \rangle$ satisfies *event commutativity* if, for all $f, g \in \mathscr{D}_1$, $f \succsim g \succsim e$, $C, D \in \mathscr{E}_{\mathbf{E}}$,

$$((f, C; g), D; g) \sim ((f, D; g), C; g). \tag{18}$$

This is a consequence of RDU, although it is much weaker than that representation. It has been sustained in several experiments involving all gains or all losses, but not in the mixed case.

Assuming that M of (13) is in class *IIIC* (Def. 1), it is immediate that event commutativity holds iff the following functional equation is satisfied:

$$M_w[M_{w'}(p, q), q] = M_{w'}[M_w(p, q), q] \qquad (w, w' \in [0, 1], 0 \leqslant q \leqslant p < k). \tag{19}$$

We shall call also this equation "event commutativity".

2.2. Uniqueness Theorem. The following result, a multiplicative version of a particular case of a uniqueness theorem reported in [9], is used to support some calculations in Theorems 1 and 4. When it is used for functions *f*, *g*, and *h* with values in $[0, \infty[$, it is often applied to the subdomain where the values are non-zero, and the concluding forms are extended by other considerations such as continuity or boundary conditions. In the proof of Theorem 1, such details are made explicit.

Theorem 0. *Let X and Y be finite or infinite real intervals*, $T : X \times Y \to \mathbb{R}$ *be continuous*, *and consider the functional equation*

$$f(x)g(y) = h(T(x,y)) \quad (x \in X, y \in Y) \tag{20}$$

where $f : X \to]0,\infty[$, $g : Y \to]0,\infty[$ *and* $h : T(X \times Y) \to]0,\infty[$. *If* (20) *has a solution* (f_0, g_0, h_0) *with continuous*, *nonconstant* f_0 *and* g_0, *then*

$$f = \beta f_0^{\alpha},\ g = \gamma g_0^{\alpha},\ h = \beta\gamma h_0^{\alpha}, \tag{21}$$

where $\beta > 0$, $\gamma > 0$, α *are arbitrary constants*, *give the general solutions* (f, g, h) *with continuous f and g*.

2.3. Characterization of *M* with Event Commutativity. We first functionally characterize event commutativity under weak hypotheses.

Theorem 1. *Suppose* $M \in IIIC$ (Def. 1). *Then M satisfies event commutativity* (19) *iff it has the representation*

$$\theta_q[M_w(p,q)] = \psi_q(w)\theta_q(p) \quad (w \in [0,1], q \in [0,k[, p \in [q,k[). \tag{22}$$

Here $\theta_q : [q,k[\to [0,\infty[$ *is strictly increasing and continuous*, $\theta_q(q) = 0$, *and* $\psi_q : [0,1] \to [0,1]$ *is an increasing bijection.*

Proof. Assume (19). Temporarily hold $q \in [0,k[$ fixed and write

$$N(w,p) := M_w(p,q) \quad (w \in [0,1], p \in [q,k[). \tag{23}$$

Then we have the functional equation

$$N[w, N(w',p)] = N[w', N(w,p)] \quad (w, w' \in [0,1], p \in [q,r]) \tag{24}$$

for each $r \in]q,k[$. We apply the following particular case of a theorem proved in [2]. If the continuous function $N : [0,1] \times [q,r] \to [q,r]$ satisfies (24) and $\{N(w,r) | w \in [0,1]\} = [q,r]$ (a consequence of continuity and of internness, (17)), then there exists a continuous commutative semigroup with identity r (monoid) $([q,r], \circ)$ such that

$$N(w,p) = N(w,r) \circ p \quad (w \in [0,1], p \in [q,r]). \tag{25}$$

In light of (16), q is the zero of the semigroup. So, by a theorem of Faucett [3] on continuous semigroups, there exists a continuous isomorphism $\phi_{r,q} : [q,r] \to [0,1]$ from the semigroup $([q,r], \circ)$ onto [0, 1] (under the usual multiplication) which is strictly increasing. With this and (23) we rewrite the representation (25) in the form

$$\phi_{r,q}(M_w(p,q)) = \phi_{r,q}(M_w(r,q))\phi_{r,q}(p) \quad (w \in [0,1], p \in [q,r]). \tag{26}$$

Considering (26) for two arbitrary r_1, r_2 in $]q,k[$ we have the pair of relations

$$\phi_{r_i,q}(M_w(p,q)) = \phi_{r_i,q}(M_w(r_i,q))\phi_{r_i,q}(p)$$
$$(i = 1,2;\ w \in [0,1],\ p \in [q,r_1] \cap [q,r_2]).$$

Before we apply the uniqueness Theorem 0 (sec. 2.2) to this equation we pay attention to where ϕ values are zero. Noticing that (i) $\phi_{r_i,q}(M_w(p,q)) = 0$ if, and only if, $M_w(p,q) = q$, (ii) $\phi_{r_i,q}(M_w(r_i,q)) = 0$ if, and only if, $M_w(r_i,q) = q$, and (iii) $\phi_{r_i,q}(p) = 0$ if, and only if, $p = q$, we stay away from $p = q$ and, cf. (16), from $w = 0$ in the application. With $X =]0,1]$, $Y =]q,r_1] \cap]q,r_2]$, $x = w$, $y = p$, $T(w,p) = M_w(p,q)$ restricted to $]0,1] \times (]q,r_1] \cap]q,r_2])$, $f_0(w) = \phi_{r_2,q}(M_w(r_2,q))$, $g_0(p) = \phi_{r_2,q}(p)$ and $h_0 = \phi_{r_2,q}(\cdot)$, $f(w) = \phi_{r_1,q}(M_w(r_1,q))$, $g(p) = \phi_{r_1,q}(p)$ and $h = \phi_{r_1,q}(\cdot)$, the uniqueness theorem implies the existence of "constants" $b > 0$, $c > 0$, d such that the three relations (21) hold:

$$\phi_{r_1,q}(M_w(r_1,q)) = b(r_1,r_2)\phi_{r_2,q}(M_w(r_2,q))^{d(r_1,r_2)} \quad (w \in]0,1]), \tag{27}$$

$$\phi_{r_1,q}(p) = c(r_1,r_2)\phi_{r_2,q}(p)^{d(r_1,r_2)} \quad (p \in]q,r_1] \cap]q,r_2]), \tag{28}$$

and

$$\phi_{r_1,q}(M_w(p,q)) = b(r_1,r_2)c(r_1,r_2)\phi_{r_2,q}(M_w(p,q))^{d(r_1,r_2)}$$

for all $w \in]0,1]$, $p \in]q,r_1] \cap]q,r_2]$. In view of internness and certainty, (15), the third relation at $w = 1$ gives $\phi_{r_1,q}(p) = b(r_1,r_2)c(r_1,r_2)\phi_{r_2,q}(p)^{d(r_1,r_2)}$, and so its consistency with (28) yields

$$b(r_1,r_2) = 1.$$

With $b(r_1,r_2) = 1$, the third relation coincides with (28) fully and can be dismissed. Extending the domains of (27) and (28) to include $w = 0$ and $p = q$ by checking the zeroes of ϕ, as mentioned in (i), (ii) and (iii), we get

$$\phi_{r_1,q}(M_w(r_1,q)) = \phi_{r_2,q}(M_w(r_2,q))^{d(r_1,r_2)} \quad (w \in [0,1]), \tag{29}$$

$$\phi_{r_1,q}(p) = c(r_1,r_2)\phi_{r_2,q}(p)^{d(r_1,r_2)} \quad (p \in [q,r_1] \cap [q,r_2]). \tag{30}$$

Because the range of $\phi_{r_i,q}$ is [0, 1], $i = 1,2$, we see that $d(r_1,r_2) > 0$. Using (30) transitively we get

$$\phi_{r_1,q}(p) = c(r_1,r_2)\phi_{r_2,q}(p)^{d(r_1,r_2)} = c(r_1,r_2)[c(r_2,r_3)\phi_{r_3,q}(p)^{d(r_2,r_3)}]^{d(r_1,r_2)}.$$

When it is compared with the more direct relation

$$\phi_{r_1,q}(p) = c(r_1,r_3)\phi_{r_3,q}(p)^{d(r_1,r_3)}$$

on the proper interval $[q,\min(r_1,r_2,r_3)]$ whose image under $\phi_{r_3,q}$ is also a proper interval, we see that d, c satisfy the connecting functional equations

$$d(r_1,r_3) = d(r_1,r_2)d(r_2,r_3), \quad c(r_1,r_3) = c(r_1,r_2)c(r_2,r_3)^{d(r_1,r_2)}.$$

Fixing an arbitrary $r_0 > q$ and defining a one-place function $d(r) := d(r_0, r)$, we get

$$d(r_1, r_2) = d(r_2)/d(r_1) \quad (r_1, r_2 \in]q, k[).$$

Thus

$$c(r_1, r_3) = c(r_1, r_2)c(r_2, r_3)^{d(r_2)/d(r_1)},$$

i.e.,

$$c(r_1, r_3)^{d(r_1)} = c(r_1, r_2)^{d(r_1)}c(r_2, r_3)^{d(r_2)}.$$

Similarly, defining a one-place function c by $c(r) := c(r_0, r)^{d(r_0)}$, we get

$$c(r_1, r_2)^{d(r_1)} = c(r_2)/c(r_1).$$

With this we simplify (29) and (30):

$$\phi_{r_1,q}(M_w(r_1, q))^{d(r_1)} = \phi_{r_2,q}(M_w(r_2, q))^{d(r_2)} \quad (w \in [0, 1]), \tag{31}$$

$$c(r_1)\phi_{r_1,q}(p)^{d(r_1)} = c(r_2)\phi_{r_2,q}(p)^{d(r_2)} \quad (p \in [q, r_1] \cap [q, r_2]). \tag{32}$$

Thus the maps ψ_q and θ_q given by

$$\psi_q(w) := \phi_{r,q}(M_w(r, q))^{d(r)} \quad (w \in [0, 1],\ r \in]q, k[), \tag{33}$$

$$\theta_q(p) := c(r)\phi_{r,q}(p)^{d(r)} \quad (p \in [q, r],\ r \in]q, k[), \tag{34}$$

are well-defined; that is, the right hand side expressions are independent of the choice of r, as conveyed by (31) and (32). The stated domains and ranges of ψ_q and θ_q are as indicated in the theorem, and the asserted monotonicity and continuity can also be observed from their definitions. Writing (26) in the form

$$\phi_{r,q}(M_w(p, q))^{d(r)} = \phi_{r,q}(M_w(r, q))^{d(r)}\phi_{r,q}(p)^{d(r)} \quad (w \in [0, 1],\ p \in [q, r])$$

and using (33) and (34) we have

$$\theta_q(M_w(p, q)) = \psi_q(w)\theta_q(p) \quad (w \in [0, 1],\ p \in [q, k[).$$

Letting q vary, this proves the representation (22).

Conversely, if M_w has the representation (22), then

$$\theta_q(M_w(M_{w'}(p, q), q)) = \psi_q(w)\psi_q(w')\theta_q(p)$$

and event commutativity (19) follows. □

2.4. Characterization under Separability and Segregation. We now add to the gamble structure the utility of terms of the form $f \oplus g$, $U(f \oplus g)$. We suppose, similarly to (12), that $U(f \oplus g)$ depends only on $U(f)$ and $U(g)$ and does so continuously in the former, that is

$$U(f \oplus g) = L[U(f), U(g)], \tag{35}$$

and so we may define L by

$$L(p, q) := U(U^{-1}(p) \oplus U^{-1}(q)) \quad (p, q \in [0, k[). \tag{36}$$

As traditional, when the second variable q of L is fixed, we denote the resulting function of the first variable by $L(\cdot, q)$. By the left monotonicity of $\oplus$, $L(\cdot, q)$ is strictly increasing for each fixed $q \in [0, k[$. The function $L(\cdot, q)$ is also assumed to map $[0, k[$ continuously onto $[q, k[$. So we can define l by

$$l(p, q) = r \Leftrightarrow L(r, q) = p \quad (0 \leqslant q \leqslant p < k). \tag{37}$$

The reason for the constraint on the arguments is that by Proposition 1 we have $p = L(r, q) \geqslant q$. It follows that $l(\cdot, q)$ maps $[q, k[$ onto $[0, k[$, and is strictly increasing and continuous. Note that

$$L[l(p, q), q] = l[L(p, q), q] = p \quad (0 \leqslant q \leqslant p < k). \tag{38}$$

A utility representation (U, W) is said to be *separable* if for all $f \in \mathscr{D}_2^+$, $C \in \mathscr{E}_{\mathbf{E}}$,

$$U(f, C; e) = U(f)W(C). \tag{39}$$

Clearly, RDU is separable.

The results in the next proposition are trivial consequences of the definitions.

Proposition 3. *Assume that $M \in IIIC$, where M is defined by* (13) *in terms of a utility representation* (U, W), *and that $\oplus$ is such that L and l are defined, respectively, by* (36) *and* (37). *Then*:

(*a*) *The element e is a left identity*, (7), *iff*

$$L(0, q) = q, \tag{40}$$

which is equivalent to

$$l(q, q) = 0. \tag{41}$$

(*b*) *Segregation* (8) *holds iff*

$$L[M_w(p, 0), q] = M_w[L(p, q), q]. \tag{42}$$

(*c*) *Separability* (39) *holds iff*

$$M_w(p, 0) = pw. \tag{43}$$

We refer to the equations (40) [or (41)], (42), and (43) as "left identity", "segregation", and "separability", respectively.

We now present a functional characterization of segregation for a separable utility representation.

Theorem 2. *Assume that $M \in IIIC$, where M is defined by* (13) *and that $\oplus$ is such that L and l are defined, respectively, by* (36) *and* (37), *in particular that $L(\cdot, q) : [0, k[\rightarrow [q, k[$ is strictly increasing and onto for all $q \in [0, k[$. Then M and L satisfy segregation* (42) *and separability* (43) *iff*

$$M_w(p, q) = L[wl(p, q), q] \quad (w \in [0, 1], 0 \leqslant q \leqslant p < k), \tag{44}$$

or, what is the same,

$$l(M_w(p, q), q) = wl(p, q), \tag{45}$$

and L satisfies the boundary condition

$$L(p,0) = cp \tag{46}$$

for some constant $c > 0$.

Proof. Suppose M and L satisfy segregation and separability. Putting the latter into the former gives

$$M_w[L(p,q),q] = L(pw,q)$$

which, by (37), is equivalent to (44) and so to (45).

Setting $q = 0$ in the above we get in particular $M_w[L(p,0),0] = L(pw,0)$. By separability, the left side is simply $wL(p,0)$ and so we have the functional equation

$$L(pw,0) = wL(p,0) \quad (w \in [0,1],\ p \in [0,k[).$$

Fixing $p = b \in]0,k[$ in the above equation and letting $c := L(b,0)/b$, we have $L(bw,0) = wL(b,0) = c(bw)$ for all $w \in [0,1]$. This shows $L(p,0) = cp$ for $p \in [0,b]$. If $p \in [0,k[$ and $p > b$, then there exists a $w \in]0,1]$ such that $pw \in]0,b]$ and so, by what we have just shown, $cpw = L(pw,0) = L(p,0)w$. Cancel the nonzero w and we also have $L(p,0) = cp$. This proves (46). Since $L(\cdot,0)$ is onto $[0,k[$, $c > 0$ follows.

Conversely, putting $q = 0$ into (44), using (46) and (37), we get

$$M_w(p,0) = cwl(p,0) = pw$$

which confirms separability. Since separability holds, $L[M_w(p,0),q] = L(wp,q)$; and, by (44) and (38),

$$M_w[L(p,q),q] = L(wl[L(p,q),q],q) = L(wp,q).$$

Thus segregation indeed follows. □

Corollary. *Suppose that the hypotheses of Theorem 2, separability, and segregation hold. Then for all* $f \in \mathscr{D}_2^+$,

$$f \oplus e \succsim f \quad \textit{iff} \ \ c \geqslant 1. \tag{47}$$

Proof. By (35) and (46)

$$U(f \oplus e) = L(U(f), U(e)) = cU(f),$$

and so $f \oplus e \succsim f$ iff $U(f \oplus e) = cU(f) \geqslant U(f)$ iff $c \geqslant 1$. □

3. Homogeneous Representations

A numerical function H of two variables is said to be *homogeneous* (of degree 1) if

$$H(\nu p, \nu q) = \nu H(p,q) \tag{48}$$

for all $\nu > 0$, provided (p,q) and $(\nu p, \nu q)$ are in the domain of H [1, p. 129]. This class of functions is widely used in the sciences because it corresponds to representations that are invariant under so-called ratio scale transformations [7, 8] (cf. [1, p. 41]). This is the usually desired level of measurement that is often achieved in the physical sciences and less often in the behavioral ones. Clearly. M_w in RDU is homogeneous.

Generally when H is homogeneous, and when $(p, 0)$ is in the domain of H, the boundary function $H(\cdot, 0)$ satisfies the equation $H(\nu p, 0) = \nu H(p, 0)$ and so is of the form $H(p, 0) = \delta p$ for some constant δ. Of course, for this to be valid, the function $H(\cdot, 0)$ has to have appropriate domain that does not contain both negative and positive numbers. A simple proof could be given similar to that of (46) in Theorem 2. With this understanding, in the subsequent representation theorems for homogeneous M_w, we will not give the detailed arguments for the linearity $M_w(p, 0) = \delta(w)p$ of the boundary function $M_w(\cdot, 0)$.

3.1. With Event Commutativity. The following shows how Theorem 1 simplifies when homogeneity is added.

Theorem 3. *Suppose $M \in IIIC$ satisfies event commutativity,* (19), *and M_w is homogeneous,* (48). *Then*

$$M_w(p, q) = q\theta^{-1}\left[\psi(w)\theta\left(\frac{p}{q}\right)\right] \quad (w \in [0, 1], 0 < q \leqslant p < k). \tag{49}$$

where $\theta : [1, \infty[\to [0, \infty[$ is strictly increasing and continuous, $\theta(1) = 0$, and $\psi : [0, 1] \to [0, 1]$ is an increasing bijection. On the boundary, $M_w(p, 0) = \delta(w)p$ for all $w \in [0, 1], 0 \leqslant p < k$ where $\delta : [0, 1] \to [0, 1]$ is an increasing bijection.

Proof. If $M \in IIIC$ and M_w is homogeneous on the domain $\{(p, q) : 0 \leqslant q \leqslant p < k\}$, then for each w, M_w always has a unique homogeneous extension $\overline{M}_w$ on the larger domain $\{(p, q) : 0 \leqslant q \leqslant p < \infty\}$, and the extended function $\overline{M}$ continues to be in the class $IIIC$ (with new $k' = \infty$) and satisfies event commutativity.

So with no loss of generality we assume $k = \infty$. By homogeneity and by the conclusion of Theorem 1, for $q > 0, p \geqslant q$,

$$\overline{M}_w(p, q) = q\overline{M}_w\left(\frac{p}{q}, 1\right) = q\theta_1^{-1}\left[\psi_1(w)\theta_1\left(\frac{p}{q}\right)\right],$$

and (49) follows by setting $\theta = \theta_1$ and $\psi = \psi_1$ and restricting the result to the original function M. On the boundary, homogeneity gives the form $M_w(p, 0) = \delta(w)p$, and because $M \in IIIC$, δ is necessarily an increasing bijection on [0, 1]. $\square$

Notice that we have in (49) $\theta(p/q)$, a special case of $\theta_q(p)$ in (22).

Corollary. *Suppose that (U, W) has the representation given by* (12), *where M is as in Theorem* 3. *Define $W^*(C) := \delta[W(C)]$ and $\chi(W^*(C)) := \psi(W(C))$, i.e. $\chi = \psi \circ \delta^{-1}$, then $\chi : [0, 1] \to [0, 1]$ is an increasing bijection and (U, W^*) has the representation*

$$U(f, C; g) = U(g)\theta^{-1}\left[\chi(W^*(C))\theta\left(\frac{U(f)}{U(g)}\right)\right] \quad (g \succ e),$$
$$U(f, C; e) = W^*(C)U(f)$$

which satisfies event commutativity, homogeneity, and separability.

3.2. Characterization of M_w under Homogeneity, Separability, and Segregation.

Theorem 4. *Assume that $M \in IIIC$ and separability* (39) *holds.*

(*a*) *If L is homogeneous, then M_w is homogeneous.*

(*b*) *Suppose that segregation* (42) *is satisfied. Then M_w is homogeneous iff there exist function $\sigma :]0,k[\to]0,\infty[$ and $\lambda : [1,\infty[\to [0,\infty[$, where λ is strictly increasing and continuous with $\lambda(1) = 0$, such that*

$$l(p,q) = \sigma(q)\lambda\left(\frac{p}{q}\right) \quad (0 < q \leqslant p < k) \tag{50}$$

and

$$L(p,q) = q\lambda^{-1}\left(\frac{p}{\sigma(q)}\right) \quad (p,q \in [0,k[, 0 < q), \tag{51}$$

in which case

$$M_w(p,q) = q\lambda^{-1}\left[w\lambda\left(\frac{p}{q}\right)\right] \quad (0 < q \leqslant p < k). \tag{52}$$

(*c*) *If segregation and homogeneity hold, then the event commutativity* (18) *also holds.*

Proof. (a) It follows from (37) and (48), that l is homogeneous iff L is homogeneous. To see this, suppose that l is homogeneous. Then, as $L(\nu r, \nu q) = \nu L(r,q) \Leftrightarrow l(\nu L(r,q), \nu q) = \nu r \Leftrightarrow \nu l(L(r,q), q) = \nu r \Leftrightarrow \nu r = \nu r$, so the homogeneity of L, $L(\nu r, \nu q) = \nu L(r,q)$, is confirmed. By symmetry, when L is homogeneous, so is l.

Using this and (44), we see

$$\begin{aligned} M_w(\nu p, \nu q) &= L[l(\nu p, \nu q)w, \nu q] \\ &= L[\nu l(p,q)w, \nu q] \\ &= \nu L[l(p,q)w, q] \\ &= \nu M_w(p,q). \end{aligned}$$

(b) Assume M_w is homogeneous and segregation holds. Representing the homogeneous M_w temporarily as $M_w(p,q) = q\varphi_w(p/q)$, we rewrite (45) as

$$l(q\varphi_w(p/q), q) = wl(p,q). \tag{53}$$

Defining

$$u := p/q, \quad f(u,q) := l(uq,q) \quad (0 < q \leqslant p < k), \tag{54}$$

equation (53) becomes the functional equation

$$f(\varphi_w(u), q) = wf(u,q).$$

Let $D_q := [1, k/q[$ and observe that $f(u,q)$ is defined for all $u \in D_q$, with $f(u,q) = 0$ only when $u = 1$. Treating $q = q_1$ and $q = q_2$ as fixed, and applying the uniqueness Theorem 0 (Sect. 2.2) with $T(w,u) := \varphi_w(u)$ on $]0,1] \times (]1,k/q_1[\cap]1,k/q_2[)$ to the two solutions $(f(\cdot,q_1), w\mapsto w)$ and $(f(\cdot,q_2), w\mapsto w)$, we see that there exist three "constants" connecting the two solutions. However, because the second components $w\mapsto w$ are the same in both pairs, two of the three constants equal one. This leaves just one constant $\sigma(q_1,q_2) > 0$ and the relation

$$f(u,q_1) = \sigma(q_1,q_2)f(u,q_2) \quad u \in D_{q_1} \cap D_{q_2}.$$

Evidently σ satisfies

$$\sigma(q,q) = 1, \quad \sigma(q_1,q_3) = \sigma(q_1,q_2)\sigma(q_2,q_3).$$

Fixing a q_0 and letting σ in a single variable be defined by $\sigma(q) = \sigma(q,q_0)$ $(q \in]0,k[)$ we have $\sigma(q_1,q_2) = \sigma(q_1)/\sigma(q_2)$. Thus

$$f(u,q_1)/\sigma(q_1) = f(u,q_2)/\sigma(q_2) \quad u \in D_{q_1} \cap D_{q_2}.$$

This equation allows us to define a function λ unambiguously on $\bigcup_{q\in]0,k[} D_q = [1,\infty[$ by

$$\lambda(u) = f(u,q)/\sigma(q) \quad (u \in D_q,\ q \in]0,k[).$$

Thus $\lambda : [1,\infty[\to [0,\infty[$. It is clear from the strict monotonicity and continuity of l, and from the definitions of f and λ, that λ is strictly increasing and continuous. Now we have

$$f(u,q) = \sigma(q)\lambda(u) \quad (u \in D_q,\ q \in]0,k[).$$

Replacing u by p/q we arrive at (50): $l(p,q) = \sigma(q)\lambda(p/q)$. We have (41) $l(q,q) = 0$, so $\lambda(1) = 0$ follows. Clearly, (50) is equivalent to (51). Putting (50) back into (45) we have

$$\lambda(M_w(p,q)/q) = w\lambda(p/q).$$

This gives the representation (52).

(c) Using (52) we have

$$M_w[M_{w'}(p,q),q] = q\lambda^{-1}\left[w\lambda\left(\frac{M_{w'}(p,q)}{q}\right)\right] = q\lambda^{-1}\left[ww'\lambda\left(\frac{p}{q}\right)\right].$$

Since this is symmetric in w, w', (19) is immediate for $q \neq 0$. Separability implies (19) for $q = 0$. Thus (18) holds. □

Since $\lambda(1) = 0$, we have $L(0,q) = q\lambda^{-1}(0) = q$, so the left-identity boundary condition is met.

Linear σ leads to homogeneous L, and thus homogeneous M_w. But, in general, σ need not be linear to generate homogeneous M_w. The function σ simply has no bearing on M, merely reflecting the non-uniqueness of the generating L.

3.3. Relation between Representations. It is important to realize that although the representation (49) of Theorem 3 looks similar to that of Theorem 4, (52), and the proofs have a common thread, the fact is that they have a distinctive aspect – we do not know when proving Theorem 4 whether or not the relation (44) among M, L, and l has a homogeneous extension. However, the following result establishes how the two representations relate.

Theorem 5. *Suppose that $M \in IIIC$ satisfies separability, segregation, and homogeneity. Then the representation* (49) *of Theorem* 3 *in terms of θ, ψ and the representation* (52) *of Theorem* 4 *in terms of λ are related as follows*: *for some constants $\rho > 0, a > 0$*

$$\psi(w) = w^\rho, \ \ \theta(z) = a\lambda(z)^\rho. \tag{55}$$

Proof. By parts (b) and (c) of Theorem 4, both of the representations (52) and (49) hold, so

$$\lambda^{-1}\left[\lambda\left(\frac{p}{q}\right)w\right] = \theta^{-1}\left[\theta\left(\frac{p}{q}\right)\psi(w)\right].$$

With $s = \psi(w) \in [0,1]$, $t = \theta(p/q)$, and $\varphi(t) := \lambda[\theta^{-1}(t)]$ this reduces to

$$\varphi(st) = \psi^{-1}(s)\varphi(t),$$

which is a Pexider equation. It can be solved directly or by applying Theorem 0 in section 2.2 of this paper. The strictly increasing solutions are $\psi^{-1}(s) = s^{\rho'}, \varphi(t) = a't^{\rho'} (a' > 0, \rho' > 0)$ [1. p. 81]. Equation (55) follows with $\rho = 1/\rho'$, $a = (a')^{-\rho}$. □

Note: Also the equation $L(pw, 0) = L(p, 0)w \, (w \in [0,1], p \in [0, k[)$ in the proof of Theorem 2 could have been solved with aid of Theorem 0. Conversely, the equations in the proofs of Theorems 1 and 4, to which we had applied Theorem 0, can be reduced to Pexider equations [1, pp. 71–81] by sequences of sometimes not very convenient substitutions.

4. Conclusions

In the context of a general utility theory of binary gambles we have explored several widely invoked properties: event commutativity, separability, segregation, and homogeneity. Various combinations of these properties were characterized by functional equations in terms of the corresponding numerical forms. Event commutativity was shown to follow from homogeneity and segregation. We intend to use some of these results in a future paper in which we study, in the context of non-commutative joint receipt and the rank-dependent utility representation, the possible forms for $U(f \oplus g)$.

References

[1] Aczél J (1987) A Short Course on Functional Equations. Dordrecht/Boston: Reidel/Kluwer

[2] Aczél J, Wallace, AD (1967) A note on generalizations of transitive systems of transformations. Colloq Math **17**: 29–34

[3] Faucett WM (1955) Compact semigroups irreducibly connected between two idempotents. Proc Amer Math Soc **6**: 741–747

[4] Luce RD (2000) Utility of gains and losses: measurement-theoretical and experimental approaches. Erlbaum, Mahwah, NJ: Erlbaum. Errata: see Luce web page at http://www.socialscience.uci.edu

[5] Luce RD (2002) A theory of psychophysical proportions, joint presentations, and matches. Psych Rev (in press)

[6] Luce RD, Fishburn PC (1991) Rank- and sign-dependent linear utility models for finite first-order gambles. J Risk Uncert **4**: 29–59

[7] Luce RD, Narens L (1985) Classification of concatenation structures by scale type. J Math Psych **29**: 1–72
[8] Narens L (1981) A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. Theory Decis **13**: 1–70
[9] Ng CT (1973) On the functional equation $f(x) + \sum_{i=1}^{n} g_i(y_i) = h[T(x, y_1, \ldots, y_n)]$. Ann Polon Math **27**: 329–336

Authors' addresses: Che Tat Ng, Department of Pure Mathematics, University of Waterloo, Waterloo, Ont. N2L 3G1, Canada, e-mail: ctng@math.uwaterloo.ca; R. Duncan Luce, Departments of Cognitive Science and Economics, University of California, Irvine, CA 92697, U.S.A., e-mail: rdluce@uci.edu; János Aczél, Department of Pure Mathematics, University of Waterloo, Waterloo, Ont. N2L 3G1, Canada, e-mail: jdaczel@math.uwaterloo.ca