ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Organizational Behavior and Human Decision Processes 89 (2002) 1176–1193

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES

www.elsevier.com/locate/obhdp

Duplex decomposition and general segregation of lotteries of a gain and a loss: An empirical evaluation ☆

Young-Hee Cho,[a,*] R. Duncan Luce,[b] and Lan Truong[a]

[a] California State University, Long Beach, CA, USA
[b] University of California, Irvine, CA, USA

Abstract

We investigate the empirical validity of two hypotheses, duplex decomposition (DD) and general segregation (GS), regarding decomposition of a binary gamble of a gain and a loss into two unitary gambles in which one consequence of each gamble is no change from the status quo. Four binary lotteries (money gambles with specified probabilities) and four decomposed lotteries designed to test GS and four decomposed lotteries to test DD were constructed, and certainty equivalents (CEs) were estimated for each lottery. Respondents' indifference between a binary lottery and a decomposed lottery was determined by evaluating the equality between the CE of a mixed binary lottery and the CE of the corresponding decomposed lottery. Given the variability of estimates of CEs and the lack of a clear statistical definition for the equality between two CEs, we applied several criteria: we counted responses where the difference between two CEs was either ±2% (the strictest criterion), ±4%, ±6%, or ±8% (the most lenient criterion) of the range of lottery outcomes. The results showed that under the strictest criterion, GS held for 25% of the responses and DD for 22%. Under the most lenient criterion, GS held for 56% of the responses and DD for 52%. Depending upon the criterion used, between 39 and 75% of the responses were consistent with at least one of the hypotheses. Several methodological problems in determining the indifference between two lotteries are discussed.

1. Introduction

For over half a century, various utility theories have been proposed to represent, in more-or-less simple numerical forms called *utility representations*, the choice behavior of individuals among risky and uncertain alternatives, i.e., gambles. These theories can be partitioned according to whether or not they distinguish between gain and loss consequences.

The earlier, classical utility theories—including expected utility (EU) theory (von Neumann & Morgenstern, 1947, and variants on it), subjective expected utility

☆ This research was supported in part by National Science Foundation Grants SBR-9808057 to the University of California, Irvine.

* Corresponding author.
E-mail address: ycho1@csulb.edu (Y.-H. Cho).

PII: S0749-5978(02)00026-2

theory (SEU) (Savage, 1954), and various rank-dependent theories (RDU) (Quiggin, 1982, 1993; Schmeidler, 1989)—did not distinguish gains from losses. These models all generate an overall bilinear representation in which each uncertain alternative, or gamble, is evaluated as a weighted sum of the utilities of the consequences associated with the several chance events. Because there is no intrinsic zero point, such representations are of interval-scale type—i.e., the choice of both unit and zero of the utility function is arbitrary. In the EU theories, the weights are the probabilities of the events occurring; in the SEU theories the weights are subjective probabilities constructed from choice behavior; and in RDU they are not finitely additive and thus are not probabilities, subjective or otherwise.

More recent theories—including prospect theory (PT) (Kahneman & Tversky, 1979), cumulative prospect theory (CPT) (Tversky & Kahneman, 1992), and rank- and sign-dependent utility (RSDU) theory (Luce, 1991, 1997, 2000; Luce & Fishburn, 1991, 1995)—do distinguish gains from losses. Making that distinction is important because many, if not most, real-world examples of gambles—such as investments, driving, flying, health decisions—are of mixed type. In these theories, there is a unique consequence, *e*, that represents *no change from the status quo*. Any consequence or gamble preferred or indifferent to *e* is called a *gain* and any less preferred, a *loss*, and different weights are used for gains and for losses. These theories have a ratio-scale representation because the utility of *e* is assumed (or forced by assumptions of the theory) to be zero. All of these authors proposed a bilinear representation for gambles with all gains and a *separate* bilinear representation for gambles with all losses. These representations are very similar in mathematical form to the earlier theories. The weights for gains add to 1 as do those for losses, but there is no necessary relation between the two types of weights.

Except for Luce (1997, 2000), all of these theories propose or derive representations for mixed gambles (i.e., those with some consequences that are gains and some that are losses) that are also bilinear in form; however, in general, the weights do not add to 1. The weights for mixed gambles use the weights obtained from gambles of pure gains and those from pure losses as follows: If an event in a mixed gamble gives rise to a gain consequence, then the gain weight is used, and if an event gives rise to a loss consequence, the loss weight is used. Since those weights are independently determined, they do not in general sum to 1.

In contrast, Luce (1997, 2000) explored alternative representations that arise from a rather different approach. In these models, pure gains and pure losses are treated bilinearly. However, the mixed gambles are decomposed into components where each component involves either just gains or just losses. Thus, the bilinear model applies to the components of just gains and of just losses, but depending upon how the separate evaluations of the components are combined, the overall representation in the mixed case may or may not be bilinear. As we note below, some empirical data suggest that the bilinear models are rejected for mixed gains and losses. In deriving representations for mixed gambles, Luce proposed two assumptions to decompose a mixed gamble into two simpler terms, one having to do with only the gain and the other having to do only with the loss. The purpose of the present study is to investigate the empirical validity of these two assumptions.

The remainder of the paper is organized into five sections. The first discusses the strategy of decomposing gambles and introduces the two specific decomposition assumptions for mixed gains and losses that are examined experimentally. The second is devoted to the experiment that attempts to evaluate their validity. The third introduces the data analysis and results. The fourth discusses these results in relation to two other studies and raises several methodological issues. And the last formulates conclusions.

2. The assumptions under study

2.1. The decomposition strategy and the joint receipt operation

To describe the strategy adopted by Luce (1997, 2000), we need some notation. Let $(x, C; y)$ denote the (binary) gamble in which a chance "experiment" is carried out and the holder of the gamble receives the consequence x if the chance event C occurs and y if it fails to occur. As a prototypic chance experiment, consider an urn with an unknown number of red and white balls and let C be a random draw of a red ball. The experimenter can reduce the degree of uncertainty by, for example, placing bounds on the number of each type of ball, with the least uncertainty being the exact number of each type. In the empirical experiment reported below, the events are described in terms of their probabilities of occurring in which case we write $(x, p; y)$, where $p = \Pr(C)$. When such a gamble has monetary consequences, as in our experiment, it is called a *lottery*. In most theories one is free to identify the pure consequence x with the lottery $(x, 1; y)$.

Let $\succsim$ denote a binary preference relation over gambles and consequences such that $g \succsim h$ means that g is preferred or indifferent to h. The converse is $g \precsim h$. Two gambles g and h are indifferent, i.e., $g \sim h$, if both $g \succsim h$ and $g \precsim h$ hold. And $g \succ h$ denotes that g is strictly preferred to h. The order $\succsim$ is assumed to be a weak order: transitive (i.e., for gambles f, g, and h, if $f \succsim g$ and $g \succsim h$ then $f \succsim h$) and connected (i.e., either $g \succsim h$ or $g \precsim h$ hold).

Luce's idea was that decision makers may attempt to decompose a general binary gamble $(x, C; y)$ either into a combination of a pure consequence and simpler subgambles such as $(x, C'; e)$ or $(e, C''; y)$, where C' and C'' are independent realizations of C, or into a combination of two subgambles. For example, let C be the event of drawing red balls as above. Then C' and C'' mean drawing red balls from either two independent or two successive plays of the gamble. The binary gambles that have e as one consequence are called *unitary*. The reason that such a reduction is desirable is that in most theories the utility of unitary gambles takes the following simple mathematical form:

$$U(x, C; e) = \begin{cases} U(x)W^+(C) & \text{if } x \succ e, \\ 0 & \text{if } x \sim e, \\ U(x)W^-(C) & \text{if } x \prec e, \end{cases} \tag{1}$$

where W^+ and W^- denote the weighting functions for gains and losses, respectively. Such representations are called *separable* (Section 3.5.2.3 of Luce, 2000).

The price of decomposing a binary gamble into two unitary gambles or into a unitary gamble and a pure consequence is that we must find a way to combine two or more unitary gambles. To this end, the concept of *joint receipt* is helpful. Joint receipt refers to the ubiquitous experience of receiving or considering more than one valued object at the same time. Examples of joint receipt in daily life are receiving several gifts on special occasions or purchasing several lotteries or stocks at once. The symbol $\oplus$ is used, in analogy to $+$, to denote formally the (partial) operation of joint receipt. The joint receipt of two sums of money x and y is denoted by $x \oplus y$. The joint receipt of two gambles is denoted by $(x, C; y) \oplus (u, D; v)$ and it is interpreted as meaning that the two underlying chance experiments are run off independently and the person holding the gambles receives one of the four possible consequence combinations: $x \oplus u$ if both C and D occur in two independent experiments, $y \oplus u$ if both $\overline{C}$, the complement of C, and D occur, and so on.

Despite the ubiquity of joint receipt in daily life, it was not formally incorporated into a mathematical derivation of a utility representation until the development of RSDU by Luce (1991) and Luce and Fishburn (1991). Kahneman and Tversky

(1979) discussed the joint receipt operation informally (see below); however, they neither formalized the concept as a mathematical operation nor invoked it in the formal statement of the PT representation.

Luce proposed that a person might use two specific possible decompositions of mixed gambles into the joint receipt of unitary gambles. They are called general segregation (GS) and duplex decomposition (DD). Using these two assumptions, he derived utility representations for mixed gambles. Before we introduce these two assumptions, we first examine the property of binary segregation, proposed by Luce and Fishburn (1991), for gambles of just gains. General segregation extends the notion of binary segregation to the realm of mixed gambles.

2.2. Binary segregation of gains

To formulate the two decomposition assumptions, we define the concept of qualitative subtraction, i.e., $\ominus$, in terms of $\oplus$. Formally, for any x and y,

$$x \ominus y \sim z \text{ if and only if } x \sim z \oplus y. \quad (2)$$

We assume that such a z exists, which is not much of an assumption when money is involved. Note that z may be, in general, either a gain when $x \succ y$ or a loss when $x \prec y$.

Consider the gamble $(x, C; y)$ with $x \succsim y \succsim e$. Suppose we subtract (qualitatively, as above) the smaller gain, y, from $(x, C; y)$ and then think of the gamble as the joint receipt of y and the modified gamble. This is *binary segregation of gains*. Formally,

$$(x, C; y) \sim (x \ominus y, C; e) \oplus y \quad (x \succsim y \succsim e), \quad (3)$$

where $(x \ominus y, C; e)$ is unitary as is the pure consequence, y, in a degenerate sense.

Cho and Luce (1995) and Cho, Luce, and von Winterfeldt (1994) studied binary segregation empirically on the assumption that $x \oplus y = x + y$ (see Luce, 2000, Section 4.4.2, for a summary). Note that we cannot replace $\oplus$ by $+$ when dealing with a joint receipt in which at least one of the terms is a gamble. The overall results supported the binary segregation assumption.

In developing their PT, Kahneman and Tversky (1979) invoked the concept of binary segregation somewhat informally in what they called an "editing" phase. They discussed only the case of monetary consequences of x and y with a known probability and assumed that the joint receipt of x and y is the same as $x + y$. Then for $x > y$, they proposed that $(x, p; y)$ could be segregated into $(x - y, p; 0)$ and y. Thus, the lottery was partitioned into a maximal riskless component, y, and the remaining risky, but unitary, component $(x - y, p; 0)$. However, neither did they explain how one would combine $U(x - y, p; 0)$ and $U(y)$ nor did they test the segregation rule empirically.

The following subsection introduces the utility representation derived from the binary segregation assumption.

2.3. Rank-dependent utility for pure gains

In what follows, we let U denote a function from the space of gambles and their joint receipt onto the positive real numbers and let W^+ denote a weighting function from events onto the interval $[0, 1]$. Using the binary segregation assumption and other assumptions (see Luce, 2000, for detail), Luce (1991) and Luce and Fishburn (1991, 1995) showed that each pair of the following properties implies that the third also holds:

(1) Binary segregation, Eq. (3).
(2) Rank-dependent utility (RDU), for all $x \succsim e, y \succsim e$,

$$U(x, C; y) = \begin{cases} U(x)W^+(C) + U(y)[1 - W^+(C)] & \text{if } x \succ y, \\ U(x) & \text{if } x \sim y, \\ U(x)[1 - W^+(\overline{C})] + U(y)W^+(\overline{C}) & \text{if } x \prec y. \end{cases} \quad (4)$$

(3) The following form for U holds over $\oplus$: for some constant δ,

$$U(x \oplus y) = U(x) + U(y) - \delta U(x)U(y), \quad (5)$$

and (U, W) is a separable representation (Eq. (1)) of unitary gambles.

The mathematical details of the derivation are in Luce (2000).

The form given in Eq. (5) is called *p-additive*, where *p* stands for polynomial. The term arises from the fact that it is the only polynomial form with a fixed 0 that can be transformed into an additive representation of $\oplus$. Clearly, when $\delta = 0$, U itself is additive. When $\delta \neq 0$, U itself is no longer additive, but by rearranging the *p*-additive form and letting $V(x) = -\frac{1}{\kappa}\ln[1 - \delta U(x)]$, where $\delta\kappa > 0$, it is not difficult to show that V is additive over $\oplus$. Rewriting U in terms of V,

$$U(x) = \frac{1}{\delta}[1 - \exp(-\kappa V(x))]. \quad (6)$$

When $\delta > 0$, $\kappa > 0$, U is called *negative exponential* and when $\delta < 0$, $\kappa < 0$, U is called *exponential.*

The following section introduces the generalization of binary segregation of gains into mixed gambles of the form $(x, C; y)$, where $x \succ e \succ y$.

2.4. General segregation of mixed gambles

The generalization, proposed by Luce (1997), is accomplished by distinguishing whether the given mixed gamble is seen as an overall gain or as an overall loss relative to e. *General* (*binary*) *segregation* is the assumption that, for $x \succsim e$, $x \succsim y$,

$$(x, C; y) \sim \begin{cases} (x \ominus y, C; e) \oplus y & \text{if } (x, C; y) \succ e, \\ e & \text{if } (x, C; y) \sim e, \\ (e, C; y \ominus x) \oplus x & \text{if } (x, C; y) \prec e. \end{cases} \quad (7)$$

In words, if $(x, C; y)$ is preferred to the status quo e, then it is perceived as indifferent to the joint receipt of y and the remaining unitary lottery after y is "subtracted" out. Note that general segregation, Eq. (7), includes binary segregation, Eq. (3), as the special case of $y \succsim e$. Thus, general segregation is the generalization of binary segregation for pure gains to the mixed case. When $(x, C; y)$ is less preferred than e, a similar decomposition is applied except that x, rather than y, is subtracted out. Again, this is the generalization of binary segregation for pure losses to the mixed case. In both cases, the mixed gamble is decomposed into the joint receipt of a unitary gamble and a pure consequence.

As an example, consider the binary lottery $(100, \frac{1}{2}; -50)$. Suppose that $x \oplus y \sim x + y$ and thus by definition of $\ominus$, $x \ominus y \sim x - y$. Also assume that 0 money represents no change from the status quo. Then GS postulates that

$$\left(100, \frac{1}{2}; -50\right) \sim \begin{cases} (150, \frac{1}{2}; 0) \oplus (-50) & \text{if } (100, \frac{1}{2}; -50) \succ 0, \\ 0 & \text{if } (100, \frac{1}{2}; -50) \sim 0, \\ (0, \frac{1}{2}; -150) \oplus 100 & \text{if } (100, \frac{1}{2}; -50) \prec 0. \end{cases}$$

In the experiment reported below, the joint receipt $(150, \frac{1}{2}; 0) \oplus (-50)$ was described to the respondents as meaning that the lottery $(150, \frac{1}{2}; 0)$ would be played and that there would be a 50% chance to win \$50 and another 50% chance to win nothing. They were also told that no matter what was received from the lottery, they would also forfeit \$50.

2.5. Duplex decomposition of mixed gambles

An alternative decomposition of mixed gambles into unitary ones is called *duplex decomposition.* It asserts that, for $x \succ e \succ y$,

$$(x, C; y) \sim (x, C'; e) \oplus (e, C''; y), \tag{8}$$

where the two gambles on the right are realized independently and C' and C'' denote the same event C in the two independent realizations of the underlying chance "experiment." Thus, DD reduces the gamble to the joint receipt of two unitary gambles, the one $(x, C'; e)$ involving just the gain and the other $(e, C''; y)$ just the loss. Unlike general segregation, DD is not normatively compelling because the gamble $(x, C; y)$ yields either x or y, but not both and not neither. Whereas, $(x, C'; e) \oplus (e, C''; y)$ yields either x, y, $x \oplus y$, or $e \oplus e \sim e$, no change from the status quo. Despite the non-rationality of the condition, we do not rule out the possibility that some decision makers may independently appraise the gain aspects from the loss ones and combine the two via joint receipt. It is, perhaps, the simplest decomposition for anyone to analyze, and simplification often prevails over correctness.

For the binary lottery $(100, \frac{1}{2}; -50)$, used as an example before, DD says that $(100, \frac{1}{2}; -50) \sim (100, \frac{1}{2}; 0) \oplus (0, \frac{1}{2}; -50)$. In the experiment below, the joint receipt of $(100, \frac{1}{2}; 0) \oplus (0, \frac{1}{2}; -50)$ was described as playing two lotteries independently and receiving one of the four combinations of consequences, \$100, \$50, \$0, or −\$50 with the prescribed probability, in this case each with probability of $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Slovic and Lichtenstein (1968); (also see Payne & Braunstein, 1971) first compared a binary lottery with its duplex decomposed lottery in an experiment designed to investigate whether preferences are affected by changes in variance when the expected value was fixed. They did not find any evidence of a difference in judged value between the original lottery and the corresponding duplex decomposed lottery, which implies that the variance of outcomes apparently is not a factor that determines preference among lotteries. Luce and Fishburn (1991) invoked this property in deriving the RSDU representation, which is a natural generalization to uncertain events of the PT that Kahneman and Tversky (1979) developed for lotteries. Cho et al. (1994) provided a further empirical test. They also did not reject DD (but see Section 6).

2.6. Utility of mixed gambles

The utility expressions for mixed gambles can be derived from the assumptions of separability and either GS or DD provided we know how $U(x \oplus y)$ is expressed in terms of $U(x)$ and $U(y)$ for $x \succsim e \succsim y$. Recall that the p-additive form was proposed for pure gains or for pure losses. For mixed consequences, Luce and Fishburn (1991) assumed that U is simply additive over $\oplus$:

$$U(x \oplus y) = U(x) + U(y) \quad (x \succsim e \succsim y).$$

From this, we can derive the representation of $U(x, C; y)$ for each of the decomposition assumptions (Luce, 2000, Chap. 6).

Luce (1997) proposed an alternative extension suggested by Eq. (6). Recall that the function V, rather than U, is additive over gains and losses separately. The generalization is to assume that additivity of V holds for mixed cases as well, i.e.,

$$V(x \oplus y) = V(x) + V(y) \quad (x \succsim e \succsim y).$$

Using the fact that $\exp(x + y) = (\exp x)(\exp y)$, one can derive expressions for $U(x, C; y)$ for the decomposition assumptions. The details are found in Chapter 7 of Luce (2000).

2.7. A study on the utility of mixed gambles

Before turning to the present experiment, we describe an earlier study that investigated the validity of utility models of mixed gambles. Luce (1997, 2000) developed the several utility models of mixed gambles using three possible forms for the utility function (i.e., power for $\delta = 0$, exponential for $\delta < 0$, and negative exponential for $\delta > 0$ as functions of money) and the two decomposition assumptions (i.e., GS and duplex decomposition). No specific form was assumed for the weighting function. Sneddon and Luce (2001) tested a subset of the models proposed by Luce. This subset of models can be partitioned into those satisfying general segregation or DD. In testing these models, Sneddon and Luce made the assumption that the function V that arose in Eq. (6) is proportional to money. Using the data ($n = 144$) reported in Cho et al. (1994), they found via parameter estimation which model best fit each individual respondent's data. When pooled together across respondents, the number of respondents whose data fit the models that are based on the DD assumption was 104 (73%), implying the superiority of DD over GS. And overall, the best fitting models accounted for a good deal of the variance—although in that respect the tests were not as powerful as one would like because not many degrees of freedom remained in the data after the parameters were estimated.

2.8. Goals of the present study

Because the validity of each model proposed by Luce depends upon the validity of the underlying assumptions, it is important to investigate these assumptions directly. Cho et al. (1994) concluded that the assumption of DD holds for the majority of respondents. Sneddon and Luce (2001) showed that the models based on DD fit data better than those based on GS. This result provides indirect implication that DD might be better than general segregation. However, no direct test concerning GS has been reported previously. The main goal of the current study is to investigate how well each of these hypotheses can explain the obtained data. Five outcomes are possible: GS but not DD can describe most of the data, DD but not GS can describe most of the data, both assumptions can equally account for most of the data, each assumption separately can account for a reasonable portion of the data and thus together, but not individually, can account for most of the data, and finally neither of the assumptions can account for the majority of data. The following section describes an experiment designed to test these assumptions.

3. Experiment

3.1. Estimating indifference

One major challenge in testing the empirical validity of general axiomatic assumptions, such as GS and DD that are stated in terms of indifference, is how to decide when a respondent is actually indifferent between two lotteries. The apparently simplest method is to ask people whether or not they are indifferent. This approach suffers from a lack of a clear criterion for saying "indifferent," and presumably the indifference criterion may vary greatly among people and over time for each individual.

The next simplest method is to ask people to choose between two alternatives. However, people's preference between two lotteries that they view as being very similar are often inconsistent over repetitions. Thus, repeated choices for the same pair are required to obtain a stable pattern of preference, and indifference is interpreted to mean that each alternative is selected approximately half the time. This procedure is not only very time-consuming, in part because the number of comparisons for n gambles is $n(n-1)$ pairs, but it has other difficulties. One is the necessary compromise between an adequate number of repetitions to establish a stable preference and the potential confounding effect of a respondent at some point recalling previous responses and repeating them. Another is what criterion to invoke as establishing "approximately half the occasions."

In an attempt to bypass the memory problem and to reduce the number of evaluations from $n(n-1)$ to n, the following strategy was pursued. One estimates the certainty equivalent (CE) for each lottery of interest and then asks whether or not the two estimated CEs are "equal" (see Birnbaum, 1997, for a summary). The *certainty equivalent* of a gamble is defined to be the amount of money for which a person is indifferent when choosing between the money and the gamble. Thus, for example, if a person feels indifferent between playing the lottery $g = (100, \frac{1}{2}; 0)$ and receiving the certain sum \$30, then \$30 is the CE of the lottery. Formally, this can be written as

$$\left(100, \tfrac{1}{2}; 0\right) \sim \mathrm{CE}\left(100, \tfrac{1}{2}; 0\right) = \$30.$$

Many studies simply asked respondents to judge the CE of a lottery, but there is considerable evidence that judged CEs are not the same as the CEs derived using choice procedures such as the one we adopted in this experiment (Birnbaum, 1997; Luce, 2000, p. 39–44).

In the current study, we estimated the CEs of lotteries using a version of a psychophysical up–down method in which the comparison money amount is systematically adjusted until it converges upon an amount approximately indifferent to the lottery. The method used, called Parameter Estimation by Sequential Testing (PEST), was first adopted for CE estimation of lotteries by Bostic, Herrnstein, and Luce (1990) and has been used in other studies (Cho & Luce, 1995; Cho et al., 1994). The basic procedure is as follows: Initially a lottery is presented with a randomly chosen sure amount of money lying within the range of the gains and losses of the lottery. A respondent indicates his or her preference by choosing either to play the lottery or to take the offered money. After a number of trials in which other lottery–money pairs are presented, the original lottery is again presented but paired with a different amount of money. This amount is systematically greater or less than the earlier amount. The size and direction of the change are determined by the respondent's previous choices. The size of the change is reduced as the process converges to the point where the respondent has difficulty in determining whether the lottery is better than the sum of money or vice versa. The presentation of the lottery is terminated when both the respondent alternates the choice (e.g., choosing the sum of money followed by choosing the lottery) and the difference in the sums of money between two consecutive trials for the same lottery is less than or equal to $\frac{1}{50}$ of the range of consequences in the lottery. The number $\frac{1}{50}$ is an experimenter-determined parameter of the PEST algorithm. The terminated lottery is replaced by a similar filler lottery that is designed to mask the removal of the original lottery and to maintain the average reappearance interval of each lottery constant throughout the experiment. The rules to determine the starting amount of money, how the money is modified as a function of behavior, and the termination criteria are collectively called a PEST algorithm. One type of PEST algorithm is described in detail in Cho and Luce (1995) and Cho et al. (1994). The only difference between the algorithm adopted in the current study and the earlier algorithm is that we permitted respondents to make seemingly irrational choices (e.g., choosing a lottery when the certain amount exceeds any consequence that can arise in the lottery). Robert Sneddon (personal communication, 2000) compared the relative biases in estimating CEs when irrational choices are and are not permitted. Computer simulations established that allowing the irrational choices results in less biased estimates of CEs than when irrational choices are excluded, and so we decided to permit irrational choices. Although allowing irrational choices seemed valid at the time of experimental design, it may have been a controversial decision (see Section 4.4).

3.2. The hypotheses examined

Although the GS and DD assumptions were stated for general gambles, the experiment we ran used only lotteries with monetary consequences and known probabilities. We make two important additional assumptions. The first is that no change from the status quo, e, is interpreted as no exchange of money and so it is operationally defined as the consequence of 0 money. This is an assumption in the sense that a decision maker, when confronted with a choice between lotteries, might establish a new status quo different from 0. For example, if the smaller consequence of one lottery is \$10 and that of the other is \$5, it seems plausible that the respondent may engage in a form of segregation by subtracting \$5 from all of the consequences and incorporating that amount into his or her status quo. This phenomenon, if it exists, is poorly understood, and no serious theory for it exists.

The second special assumption we make is that $x \oplus y \sim x + y$. It is an assumption that definitely may not be valid (see Luce, 2000, Section 4.5.1; Thaler, 1985). On the other hand, Cho and Luce (1995) did not reject $\mathrm{CE}(x \oplus y) = \mathrm{CE}(x + y)$ when CEs were determined by a PEST algorithm similar to the one we used in the present experiment. The result suggests that $x \oplus y \sim x + y$ is a reasonable assumption in the context of the current experimental framework.

In terms of CEs, the hypotheses to be tested take the following form. General segregation: for $x > 0, y > 0$,

$$\mathrm{CE}(x,p;-y) = \begin{cases} \mathrm{CE}[(x+y,p;0) \oplus (-y)] & \text{if } \mathrm{CE}(x,p;-y) \geqslant 0, \\ \mathrm{CE}[(0,p;-x-y) \oplus x] & \text{if } \mathrm{CE}(x,p;-y) < 0. \end{cases} \tag{9}$$

DD takes the form

$$\mathrm{CE}(x,p;-y) = \mathrm{CE}[(x,p;0) \oplus (0,p;-y)]. \tag{10}$$

4. Methods

4.1. Participants

Seventy undergraduate students at the California State University at Long Beach participated in the experiment in exchange for extra credit for the courses in which they were enrolled. All completed the experiment, but as is explained under Section 4.4, 14 were omitted from further analysis on the basis of inappropriate responses. So the results are based on the data from 56 respondents.

4.2. Stimuli

Four mixed binary lotteries and their DD and GS variants were constructed. The expected values of two binary lotteries were positive and those for the other two were negative. The four binary lotteries and their EVs are presented in Table 1.

Hereafter, we deal with estimated CEs, which we denote by $\widehat{\mathrm{CE}}$. Recall that there are two versions of GS depending upon whether a lottery is preferred to the status quo or not, as shown in Eq. (9). Thus, ideally, we should first estimate $\mathrm{CE}(x,p;-y)$ and then estimate either $\mathrm{CE}[(x+y,p;0) \oplus (-y)]$ if $\widehat{\mathrm{CE}}(x,p;-y) > 0$ or $\mathrm{CE}[(0,p;-x-y) \oplus x]$ if $\widehat{\mathrm{CE}}(x,p;-y) < 0$. To do so would require two separate sessions, which we wished to avoid. Instead, we estimated both $\mathrm{CE}[(x+y,p;0) \oplus (-y)]$ and $\mathrm{CE}[(0,p;-x-y) \oplus x]$ as well as $\mathrm{CE}(x,p;-y)$ in one session, and used whichever is

Table 1
Binary lotteries used in the experiment

Lottery	EV
(96, .8; −40)	68.8
(29, .5; −170)	−70.5
(88, .9; −140)	65.2
(13, .7; −260)	−68.9

appropriate depending on the sign of $\widehat{\mathrm{CE}}(x, p; -y)$ and ignored the other estimate. Thus, for the four binary lotteries shown in Table 1, four duplex-decomposed lotteries and eight segregated lotteries were constructed. In addition, filler lotteries were constructed by multiplying the outcomes of the test lotteries by 1/3, 1/2, 2, or 3, resulting in a total of $16 \times 4 = 64$ filler lotteries.

4.3. Procedures

The experiment was run individually. At the beginning of the practice trials, an experimenter provided instructions to a respondent using a binary lottery and a sum of money presented on a computer screen (Fig. 1). The experimenter explained the meaning of the binary lottery by demonstrating an actual chance spinner and explained that the respondent's task was to choose either the lottery or the sum of money according to their preference and to click the appropriate 'OK' button presented at the bottom of the screen. The respondent first practiced three trials of choosing between gain binary lotteries and sums of money while the experimenter was watching. The experimenter monitored the respondent's choices and provided any necessary clarification. Next, the respondent did three practice trials with loss binary lotteries followed by 3 more with mixed binary lotteries. After practicing the binary lotteries, the computer presented the typical display of the joint receipt of a lottery and a sum of money used to test the GS assumption (Fig. 2) where "&"

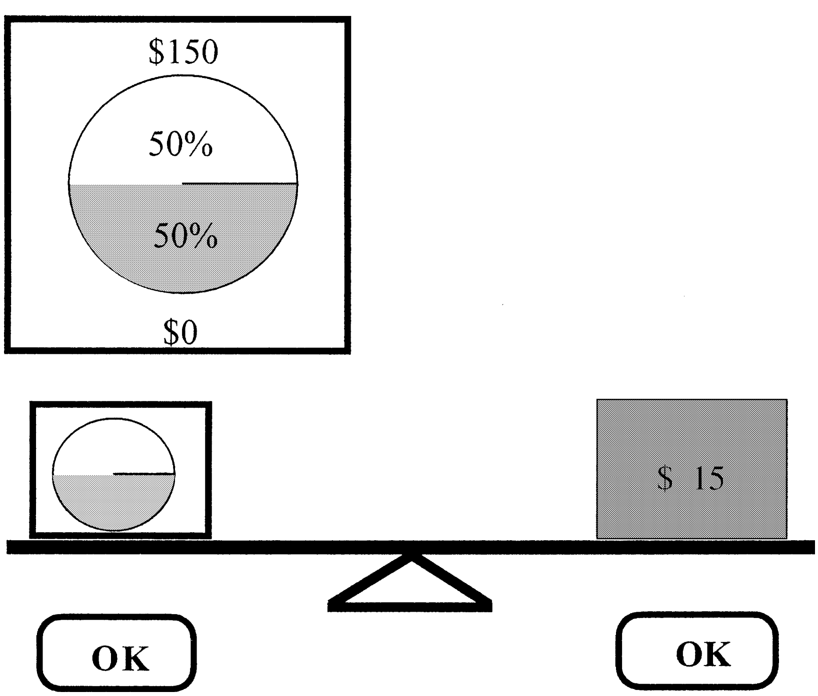

Fig. 1. A typical display of a binary gamble used in estimating a certainty equivalent of a binary gamble.

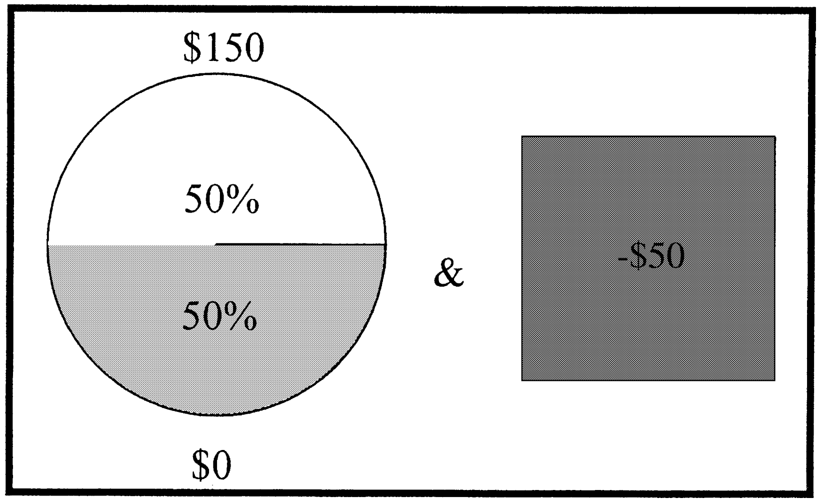

Fig. 2. A typical display of the joint receipt of a lottery and a sum of money.

represents the joint receipt of two objects. The experimenter explained to the respondent the concept of the joint receipt of two lotteries and illustrated all of the possible outcomes with their corresponding probabilities. The respondent practiced GS displays presented in a similar way as those of binary lotteries. Practice trials for the joint receipt of two lotteries used to test the duplex decomposition assumption (Fig. 3) were also given. The experiment was self-paced and the experimenter remained in the background. A set of 16 lotteries was presented in a random order. Respondents were encouraged to take a break every half an hour. The experiment was completed when all 16 of the test lotteries were replaced by the filler lotteries. The completion of the experiment took an average of 324 (from 240 to 624) trials which took an average of 43 (from 17 to 90) min per respondent.

4.4. Data analysis

The estimated CEs were first examined to determine whether they were inappropriate in the sense of lying outside the range of consequences in the lottery—either larger than the maximum consequence of the given lottery or smaller than the

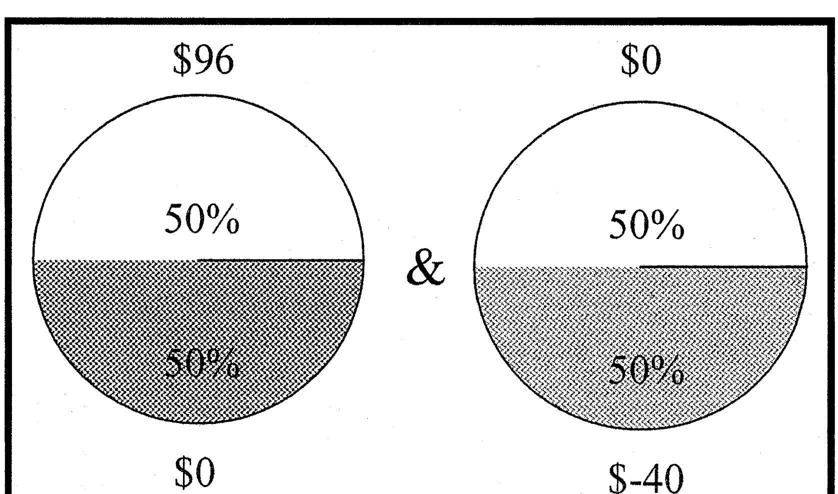

Fig. 3. A typical display for the joint receipt of two unitary lotteries.

minimum one. For example, a $\widehat{\mathrm{CE}}$ of either \$110 or −\$50 was considered inappropriate for the lottery (96, 0.8; −40). The 14 respondents who produced at least one inappropriate $\widehat{\mathrm{CE}}$ were eliminated from the data analysis. We felt that those respondents did not pay sufficiently close attention to the task to be retained. (For an alternative possibility, see the Discussion section.) Thus, the results reported are based on the data from the remaining 56 respondents.

The main focus of the data analysis was to test for each individual the equality between the estimated CE for a binary lottery, $\widehat{\mathrm{CE}}_{\mathrm{b}}$, and either the estimated CE of the decomposed version of the lottery, $\widehat{\mathrm{CE}}_{\mathrm{d}}$, or the estimated CE of the segregated version of the lottery, $\widehat{\mathrm{CE}}_{\mathrm{s}}$.

How to judge the "equality" of two estimated CEs was an issue. How different should two $\widehat{\mathrm{CE}}$s be to conclude that the hypothesis of indifference between the corresponding lotteries is rejected? It seems unreasonable to conclude that two lotteries are different when their CEs are very close, albeit numerically different (e.g., \$100 and \$101). To gauge the degree within which we can assume the relative equality of two $\widehat{\mathrm{CE}}$s, we drew upon the stopping criterion of the PEST procedure. Recall that the PEST procedure adopts a specific stopping criterion to derive each $\widehat{\mathrm{CE}}$. Specifically, in our case, when the difference between the two sure amounts of money in two consecutive trials became less than or equal to $\frac{1}{50}$, or 2%, of the range of consequences of the lottery, then the $\widehat{\mathrm{CE}}$ of the lottery ($\widehat{\mathrm{CE}}_{\mathrm{b}}$, $\widehat{\mathrm{CE}}_{\mathrm{d}}$, or $\widehat{\mathrm{CE}}_{\mathrm{s}}$) was computed as the average of these two money amounts. As noted earlier, $\frac{1}{50}$ is a parameter of the PEST algorithm. For example, for a given lottery of $(100, \frac{1}{2}; -10)$, the stopping criterion is $(100 + 10)/50 = \$2.2$. If a person chooses the money amount when given a choice between the lottery and \$47 and chooses the lottery in the next trial given a choice between the lottery and \$45, then the $\widehat{\mathrm{CE}}$ of the lottery is \$46. This algorithm treats a respondent as being indifferent in choosing between the given lottery and any sure amount of money when it lies within the stopping criterion, i.e., a money amount between \$45 and \$47.

Let S_i, where $i = \mathrm{b}$, d, or s, denote the stopping criterion for a binary lottery, a duplex decomposed lottery, and a segregated lottery, respectively. We define $I = \frac{1}{2}(S_{\mathrm{b}} + S_i)$, $i = \mathrm{d}$ or s. We explain in detail the data analysis for the test of GS; that for DD is similar. Then, $\widehat{\mathrm{CE}}_{\mathrm{b}}$ and $\widehat{\mathrm{CE}}_{\mathrm{s}}$ are treated as "equal at level I" if the difference between the two $\widehat{\mathrm{CE}}$s is less than I. Or put another way, we define $\widehat{\mathrm{CE}}_{\mathrm{b}} =_I \widehat{\mathrm{CE}}_{\mathrm{s}}$ if $|\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}}| \leqslant I$ and $\widehat{\mathrm{CE}}_{\mathrm{b}} \neq_I \widehat{\mathrm{CE}}_{\mathrm{s}}$ if $|\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}}| > I$. Of course, demanding this level of accuracy, which is ±2% of the average range of possible consequences, is quite stringent. To see what happens when the size of the interval is increased, we counted responses that fall into the range of $|\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}}| \leqslant nI$, i.e., $-nI < \widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}} < nI$, as we increases n. To do so, we first partitioned the data into 12 intervals of

$$< -5I, -5I, -4I, -3I, -2I, -I, I, 2I, 3I, 4I, 5I, > 5I,$$

where for GS the interval $-2I$, for example, means that the observed difference $\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}}$ falls in $-2I \leqslant \widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}} < -I$ whereas the interval identified as $2I$ means that $I \leqslant \widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}} < 2I$. Thus, for example, the number of responses that fall into the range of $|\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{s}}| \leqslant 2I$ is the sum of all of the responses that are in the intervals of $-2I$, $-I$, I, and $2I$. The data to test DD were analyzed in a similar way.

5. Results

5.1. Lottery data

Table 2 shows the frequency distribution of responses as a function of indifference interval I as described above. To make the table as compact as possible, we show only values of n corresponding to the interval nI described above. There were a total

Table 2
The frequency distribution of CE differences, d, as a function of indifference interval I defined in the text for GS and DD

DD	GS												$\sum$
	<-5	-5	-4	-3	-2	-1	1	2	3	4	5	>5	
<-5	10	3	1	4	0	1	4	1	0	1	2	4	31
-5	0	0	0	1	0	1	0	1	0	0	1	0	4
-4	2	1	0	0	0	0	1	0	0	0	1	1	6
-3	1	2	1	0	0	0	0	1	2	0	0	0	7
-2	5	2	0	3	1	1	2	1	0	0	0	0	15
-1	3	0	2	0	1	2	7	1	1	1	1	1	20
1	7	1	0	2	1	3	5	3	1	3	1	2	29
2	3	1	1	0	0	1	5	2	2	1	0	1	17
3	1	0	1	1	1	2	0	3	1	0	1	2	13
4	0	0	0	0	0	2	1	0	2	0	1	3	9
5	1	1	0	0	1	1	1	0	1	2	0	4	12
>5	6	0	3	2	1	5	10	2	2	7	2	21	61
$\sum$	39	11	9	13	6	19	36	15	12	15	10	39	224

of 224 responses (four pairs of $\widehat{\mathrm{CE}}_\mathrm{b} - \widehat{\mathrm{CE}}_\mathrm{s} \times 56$ respondents) for the test of GS. The rows of Table 2 show the frequency distribution of these responses. Also, there were a total of 224 responses (4 pairs of $\widehat{\mathrm{CE}}_\mathrm{b} - \widehat{\mathrm{CE}}_\mathrm{d} \times 56$ respondents) for the test of DD. The columns of Table 2 show the frequency distribution of these responses. The totals in rows and columns provide the number of responses for GS and DD, respectively. The full matrix allows the reader to investigate the number of responses that met either GS only, DD only, or both at any level of accuracy. For example, the numbers in category 3 show that 12 responses held for GS only, 13 for DD only, and only one response held for both assumptions.

At the strictest criterion, $n = 1$, i.e., $\pm 2\%$ of the range of consequences in the lottery, we obtained the following pattern. General segregation held for 25% $[= (\frac{19+36}{224}) \times 100]$ of the responses and failed for 75% of the responses. Duplex decomposition held for 22% $[= (\frac{20+29}{224}) \times 100]$ of the responses and failed for 78% of the responses. And for 39% of the responses, either GS or DD or both held. This means that the odds are about 3–1 of more unequal responses than equal responses for either hypothesis and about 3–2 against at least one of them holding. The results imply that neither assumption holds very well at the strictest level of equality interval.

Table 3 shows what happens as the equality interval is broadened, i.e., n is increased. If we accept the $\pm 4\%$ or $n = 2$ criterion as plausible, then 54% of the data support one or the other or both. But clearly something else is happening for the other 46% of the responses. Observe also in Table 3 that at all levels, neither general segregation nor DD is favored over the other, and that beyond the $\pm 2\%$ level the proportion of only GS or only DD is constant at about 20% independent of the level. The "both" category increases nearly linearly. Although the "both" category must

Table 3
Percentage of cases in Table 2 for which GS, DD, and either or both were satisfied at accuracy levels of $\pm 2n\%$, where $n = 1, 2, 3, 4$

	Indifference interval			
	±2%	±4%	±6%	±8%
GS only	17	18	20	24
DD only	14	20	20	20
Both	8	16	25	32
Neither	61	46	35	25

Table 4
Percentage of respondents exhibiting the several types of consistency over the four lotteries

Consistency	Indifference interval			
	±2%	±4%	±6%	±8%
C	13	27	45	55
I	39	46	43	39
N	48	27	13	5

Note. C = responses that are consistent with at least one of the hypotheses; N = responses for which neither of the hypotheses holds consistently; I = inconsistent responses.

increase as the criterion is relaxed, it is neither obvious that the increase should be linear nor that the GS only and the DD only categories should stay approximately constant.

5.2. Consistency of respondents

We also analyzed the data to find out the degree of response consistency of each respondent across the different stimuli. Recall that the general segregation and DD assumptions were each tested using four stimuli. Let $(\mathrm{GS}_i, \mathrm{DD}_j)$ be any pattern in which general segregation is satisfied for i stimuli and DD is satisfied for j stimuli, where $i, j = 0, 1, 2, 3, 4$. For example, $(\mathrm{GS}_4, \mathrm{DD}_4)$ means that for each respondent, both GS and DD were satisfied for all 4 stimuli of both GS and DD whereas $(\mathrm{GS}_3, \mathrm{DD}_0)$ means that GS was satisfied for three stimuli, but DD was not satisfied for any of the four stimuli. We classified an individual as consistent across stimuli when either GS or DD was satisfied for at least 3 out of 4 of the stimuli, i.e., either $i = 3, 4$ or $j = 3, 4$ or both. The percentage of these individuals is given in Table 4 under the category of 'C' meaning consistent with at least one of the hypotheses. Likewise, we counted the number of individuals whose responses showed that neither GS nor duplex decomposition was satisfied for at least 3 out of 4 stimuli, i.e., the cases where $i + j = 0, 1$. The percentage is given in Table 4 under the category of 'N' meaning neither of the hypotheses holding consistently. The remaining cases were viewed as inconsistent and are listed under the category of 'I.'

The results showed that, independent of the accuracy levels, slightly more than 40% of the respondents were inconsistent over the lotteries. The other 60% were consistent, with N predominating (48%) at the strictest accuracy level and C predominating (55%) at the least strict level.

6. Discussion

Depending upon the criterion one uses for accepting the "equality" of two certainty equivalents, one concludes that somewhere between 25 and 61% of the responses fail to support either GS or DD. The number of responses satisfying the GS assumption only is about the same as the number satisfying the DD assumption only. These findings make clear that different people are handling mixed binary gambles differently. Some appear to treat a mixed binary gamble indifferently to the joint receipt of two unitary gambles, and others appear to treat the gamble indifferently to the joint receipt of a fixed sum and a unitary gamble. But a substantial fraction of our respondents did not exhibit either pattern. Evidently, the results imply that a binary mixed gamble is not equivalents to the joint receipt of the decomposed unitary gambles. We are, therefore, in need of a better theoretical analysis for the mixed gambles.

It is difficult to predict whether some other forms of decomposition in addition to DD or GS are needed or whether a quite different axiomatic approach is needed.

This is a challenge for the theorist. Two possible alternative decompositions are to modify the current GS assumption. Instead of segregating a different amount of money depending upon whether $CE(x, p; -y)$ is positive or negative, we could invoke either

$$\mathrm{CE}(x, p; -y) = \mathrm{CE}[(x + y, p; 0) \oplus (-y)]$$

or

$$\mathrm{CE}(x, p; -y) = \mathrm{CE}[(0, p; -x - y) \oplus x]$$

across all lotteries. This has not been attempted.

We compare our findings from the present study with two other studies that, on the surface, seem to disagree with these results.

6.1. Relation to Cho et al. (1994)

The results of the current study appear to be inconsistent with the data analysis given in Cho et al. (1994) that was interpreted as supporting DD. (GS had not been proposed at the time.) This conclusion was based on a failure to reject the null hypothesis using the median test. Cho et al. employed six stimuli with two sets of money consequences, $(96, p; -40)$ and $(96, p; -160)$, and three probabilities $(p = .2, .5, .8)$ under each set of consequences. We reanalyzed their results using the current method. The results are shown in Table 5. Because there was no difference in response pattern across lotteries, the responses were pooled over the lotteries. The results showed that the DD assumption held for 24, 35, 42, and 50% of the responses for ±2, ±4, ±6, and ±8% of the outcome range, respectively. These results are virtually identical with 22, 36, 45, and 52% of the present study (sum of DD only and both rows of Table 3). This reanalysis means that the studies really agree and that the DD assumption holds only for a fraction of respondents.

The primary reason that the new analysis led to the opposite conclusion from the previous analysis is that the median test simply asks whether the distribution of $\widehat{\mathrm{CE}}_{\mathrm{b}} - \widehat{\mathrm{CE}}_{\mathrm{d}}$ is symmetric or not. Accepting the null hypothesis means that it is symmetric. However, the symmetry of a distribution does not tell us whether it is narrow or broad or even bimodal. Thus, it seems that Cho et al.'s conclusion that the DD assumption was supported was the result of a specious assumption regarding the symmetry of the distribution of responses. As the current data analytic method also has some problems as we will discuss below, a final conclusion requires further studies.

6.2. Relation to Sneddon and Luce (2001)

The conclusion of the current study, namely, that GS and DD hold in about the same fraction of cases, seems, at first, to disagree sharply with a conclusion of Sneddon and Luce (2001). As described earlier, they used a subset of the data collected by Cho et al. (1994) to evaluate comparatively several models proposed by Luce (1997, 2000). Half of the models were based on GS and the other half on DD. They explored power, exponential, and negative exponential utility functions, each of which introduced some parameters, and treated the weights as additional pa-

Table 5
Percentage of cases in Cho et al. (1994) in which DD was satisfied at accuracy levels $\pm 2n\%$, $n = 1, 2, 3, 4$

	Indifference level			
	±2%	±4%	±6%	±8%
DD:	24	35	42	50

rameters (i.e., they did not assume any specific mathematical form for the weighting function). Estimating parameters for the best fits of each model, they found that those based on DD fit the data better than those based on general segregation in a ratio of slightly less than 3–1. However, Sneddon and Luce asked a different question from ours. Specifically, on the assumption that one of the two decompositions held, they investigated which assumption led to a better fitting representation. In contrast, the present study investigated the validity of each of these assumptions rather than the comparative superiority of one assumption over another. Because the models examined by Sneddon and Luce did not admit the possibility of neither holding, the result of the current study cannot be directly compared to theirs. The comparable analysis in our data would be to ask which of DD and GS gives the better fit for the Both and Neither categories. However, finding the relative superiority between GS and DD with Neither category seems useless. This does not mean, however, that it is useless to ascertain, as Sneddon and Luce did, how well the various types of representations fit the data. Their finding that assumptions leading to non-bilinear representations are to be favored over those leading to bilinear ones, may provide some insights into the direction of future modeling.

6.3. Methodological issues in evaluating indifferences

The major challenge in testing the empirical validity of axiomatic assumptions, such as GS and DD, that assert indifference between gambles or lotteries is how to establish indifference between lotteries. This challenge raises two related questions which we discussed earlier: Should we compare two lotteries directly or indirectly? For any choice of strategy, how best do we evaluate the results statistically to determine indifference?

As was discussed in Section 3, we opted to compare lotteries indirectly by estimating certainty equivalents. It was to avoid a difficulty associated with comparing two lotteries repeatedly, namely, that respondents may ultimately remember a previous choice and simply repeat it without further consideration. To estimate the CEs of lotteries we adopted a PEST algorithm in which each lottery is paired with sums of money that are constantly adjusted according to the history of previous responses. Although the PEST procedure eliminated the memory problem associated with repeated choices, it did not eliminate the problem of how to establish the equality of two estimated CEs. If we obtained multiple CEs for each lottery, the equality of two estimated CEs could be determined statistically. However, with the PEST procedure where each estimate requires an average of 12–15 choices and so estimating 12 CEs can take up to 2 h, obtaining multiple CEs for each lottery is not easy to implement.

Due to the time constraint, we obtained only one CE for each lottery and have used the stopping criterion of the PEST procedure as a way to define the equality of two CEs. However, this approach may be flawed for two reasons. One is that we really do not understand how the stopping criterion is related to variability in the CE estimates. Specifically, it is possible that the stopping criterion may be satisfied prematurely in which case the obtained CE can be quite different from the true CE. Evidence suggesting that the premature termination may have occurred is the fact that for 14 respondents one or more CEs fell outside the range of the consequences of the lottery being estimated, which prompted us to discard their data. We did not force respondents to make reasonable choices because computer simulations (R. Sneddon, personal communication, 2000) showed that permitting irrational choices results in CE estimates that are less biased than when irrational choices are not allowed. Although simple inattention might have caused this result, we also cannot rule out the possibility that responses are probabilistically determined and so with some small probability our criterion for terminating PEST is fulfilled when the money amount is not close to the true CE of the respondent. We call this *premature* *termination.* If only one CE is prematurely determined, the two estimated CEs could be quite unequal. Put another way, such premature terminations may mean that the CE estimates are highly variable. In psychophysics where PEST procedures were initially developed, one does not stop with as few reversals as we have used. Instead, several additional reversals are usually required after meeting the first stopping criterion and the final CE was the average of last several reversal points. Again, due to the general slowness of the PEST procedure when working with gambles as compared with simple psychophysical signals, we terminated the PEST procedure when the stopping criterion was first met. Maybe that was a false economy.

A second concern is that the stopping criterion is arbitrary. Why 1/50 of the range rather than 1/100 or 1/20? Our choice was again based on a compromise between time and accuracy. Thus, it is important to find a better justified criterion for the equality of two CEs. It may be possible that estimating multiple CEs for each lottery from each participant and using the variability of CEs as a gauge for an indifference interval would mitigate the uncertainty associated with the size of the stopping criterion.

If premature terminations really do occur, then such repeated estimates of the CE of a gamble confound the variability of estimates of the actual CE with that of premature terminations. Thus, such an error estimate would be based on a mixture of distributions with quite different parameters. The presumably infrequent premature terminations will be outliers relative to the distribution of interest, but will dominate the variance estimates. Simulations are clearly needed to explore just how serious premature terminations may prove to be. We do not believe that the conclusion will depend very significantly on the particular form chosen for the psychometric function so long as it is strictly increasing, reasonably symmetric, and locally linear.

Another, closely related, challenge is to decide what statistics are appropriate for testing indifference between two lotteries. Proving the validity of axiomatic assumptions that are stated in terms of equality between two CEs requires that one accepts the null hypothesis of equality. This is antithetical to much philosophy of statistical testing developed in the 20th century and that is used widely in all of psychology. For example, in the current study, the null hypothesis for general segregation states that $CE_b = CE_s$. In a study aimed at rejecting the null, one might use a simple t-test with repeated measures. However, if the data show a clear bimodal distribution with $CE_b > CE_s$ for half of the responses and $CE_b < CE_s$ for the remaining half, then GS may not be rejected because of a large standard deviation. This danger makes it difficult to find an appropriate statistic for testing the validity of axiomatic assumptions. This danger also may be intensified when we have more than one lottery pair, as in the current study, and these are pooled. Such pooling may result in a larger estimated variance than is true for the unpooled data, thereby making it easier to retain the null hypothesis. For example, suppose that one respondent has a distribution that is skewed to the right and another respondent has the mirror image distribution that is skewed to the left. A mix of them has a symmetric distribution with the two long tails arising from the skewness, thus leading to a larger variance than that of either distribution alone. This is the reason that it is important to try to classify respondents into types before pooling over them. On the other hand, sufficiently large samples will also make it easier to reject null hypothesis.

7. Conclusions

If the PEST procedure we used provides valid estimates of CEs and if our arbitrary method of using the stopping criterion for determining the equality of two certainly equivalents is accepted, then we can conclude that neither of the hypotheses alone can hold for majority of data because either hypothesis can only hold for half of the responses even with the ±8% criterion of determining equality. However, both hypotheses together can hold for about 75% of the responses with the ±8% criterion. With the narrowest criterion of ±2%, the two hypotheses together account for only 39% of the responses. The unsolved questions regarding the validity of estimated CEs and the plausibility of the criterion for evaluating equality of estimates mean that this conclusion should be accepted with caution. Because the hypotheses of GS and DD are fundamental to the utility representations of mixed gambles developed by Luce (1997, 2000), further studies to develop a better methodology are needed to evaluate the proposed utility representations.

References

Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Lawrence Erlbaum.

Bostic, R., Herrnstein, R. J., & Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior and Organization, 13*, 193–212.

Cho, Y., & Luce, R. D. (1995). Tests of hypotheses about certainty equivalents and joint receipt of gambles. *Organization Behavior and Human Decision Processes, 64*, 229–248.

Cho, Y., Luce, R. D., & von Winterfeldt, D. (1994). Tests of assumptions about the joint receipt of gambles in rank- and sign-dependent utility theory. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 931–943.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Luce, R. D. (1991). Rank- and sign-dependent linear utility models for binary gambles. *Journal of Economic Theory, 53*, 75–100.

Luce, R. D. (1997). Associative joint receipts. *Mathematical Social Sciences, 34*, 51–74.

Luce, R. D. (2000). *Utility of uncertain gains and losses: Measurement-theoretic and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum, See Luce web site. Available: www.socsci.uci.edu, for an errata.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty, 4*, 29–59.

Luce, R. D., & Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty, 11*, 5–16.

Payne, J. W., & Braunstein, M. L. (1971). Preferences among gambles with equal underlying distributions. *Journal of Experimental Psychology, 87*, 13–18.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3*, 323–343.

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model*. Boston, MA: Kluwer Academic Publishers.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica, 57*, 571–587.

Slovic, P., & Lichtenstein, S. (1968). Importance of variance preferences in gambling decisions. *Journal of Experimental Psychology, 78*, 646–654.

Sneddon, R., & Luce, R. D. (2001). Empirical comparisons of bilinear and non-bilinear utility theories. *Organizational Behavior and Human Decision Processes, 84*, 71–94.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science, 36*, 199–214.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Received 25 June 2001