ECONOMIC RESEARCH REPORTS

THE BOX PROBLEM:
TO SWITCH OR NOT TO SWITCH

BY

Steven J. Brams

and

D. Marc Kilgour

R.R. #90-26

July, 1990

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

THE BOX PROBLEM: TO SWITCH OR NOT TO SWITCH

Steven J. Brams
Department of Politics
New York University
New York, NY 10003
USA

D. Marc Kilgour
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
CANADA

ABSTRACT

THE BOX PROBLEM: TO SWITCH OR NOT TO SWITCH

A person P is shown two identical boxes, one of which contains twice as much money as the other. P must choose one box and, after opening it, decide whether to keep this box of exchange it for the other one.

The box problem is that a switch would appear always to be profitable, no matter what the box initially chosen contains. However, when a priori probability distributions over the amounts in the two boxes are specified, this is not always the case. Nevertheless, there are both discrete and continuous probability distributions in which a switch is always rational. Moreover, whatever the distribution, a switch is rational if the amount found in the box initially chosen is sufficiently small. Implications of these findings for determining when envy is justified are discussed; also, the relationship of the box problem to the St. Petersburg paradox is briefly explored.

THE BOX PROBLEM: TO SWITCH OR NOT TO SWITCH

Steven J. Brams, New York University
D. Marc Kilgour, Wilfrid Laurier University

1. Introduction

Consider the following situation. A person P is shown two identical boxes, one of which contains \$x and the other of which contains \$2x. P must pick one box and, after opening it, decide whether to keep this box or exchange it for the other one.

Let B1 be the box that P picks initially and B2 the other box. Assume B1 contains \$100, so B2 contains either \$50 or \$200, each with probability 1/2. It would appear that P's expected value from exchanging is

$$(1/2)(\$50) + (1/2)(\$200) = \$125,$$

so P should make the switch, given that P's objective is to maximize expected value.

What if P had picked B2 initially? P's expected value from switching to B1 would also be 25 percent greater than the amount in B2, <u>so P should switch whichever box is picked initially</u>. But this is a paradox: P cannot do better, even on an expected-value basis, by switching from B1 to B2 <u>and</u> by switching from B2 to B1.

In fact, the flaw in the switching argument is that it presupposes information not given in the statement of the box problem. The information needed to determine whether or not switching is worthwhile is the a priori probability distribution of the amount of money to be found in each box. This distribution enables P to calculate whether the amount found in the box picked initially is less than the expected value of what is in the other box--and, therefore, whether a switch is profitable.

We shall show that a switch is justified if and only if (iff) the conditional probability that B1 contains the larger amount, once its contents are observed, is less than 2/3. Surprisingly, there are both discrete and continuous probability distributions for which this condition never fails: a switch is <u>always</u> called for, no matter how large the amount in B1. If a switch is not always called for, there is a threshold, below which P should switch (a switch may be rational at other times as well). Hence, for at least some amounts in B1, switching is rational.

Whether switching is occasionally or always rational depends on the a priori distribution assumed. We compare this result with the St. Petersburg paradox and discuss its relevance to explaining the rational basis of envy and risk-taking behavior.

<u>2. The Exchange Condition</u>

Let L be a random variable representing the larger amount in the two boxes, and let S represent the smaller amount. Because S = L/2, an a priori distribution on L defines an a priori distribution on S such that

$$Pr(L \le x) = Pr(S \le x/2),$$

where $0 < x < \infty$. Hence, each box always contains some positive amount.

Suppose that P picks one of the two boxes at random. Call this box B1, and denote its contents by the random variable X. Then

$$Pr(X = L) = Pr(X = S) = 1/2.$$

Now let P observe the contents of B1, X = x. A switch will be profitable for P iff the expected value of the amount in the other box, B2, is greater than x:

$$(x/2)\Pr(X = L \mid X = x) + (2x)\Pr(X = S \mid X = x) > x$$
$$(1/2)\Pr(X = L \mid X = x) + (2)\Pr(X = S \mid X = x) > 1. \quad (1)$$

If B1 is not the L box, it must be the S box, so

$$\Pr(X = S \mid X = x) = 1 - \Pr(X = L \mid X = x). \quad (2)$$

Substituting (2) into (1),

$$(1/2)(\Pr(X = L \mid X = x) + (2)[1 - \Pr(X = L \mid X = x)] > 1.$$
$$(3/2)\Pr(X = L \mid X = x) > 1,$$

so an exchange is profitable for P iff

$$\Pr(X = L \mid X = x) < 2/3. \quad (3)$$

P should switch, therefore, iff B1 has a conditional probability of less than 2/3 of being the L box, based on the observed X = x and the assumed probability distribution.

3. Discrete Distributions

We first consider the case when X, L, and S are discrete random variables. Assume there exists a fixed m > 0 such that

$$\Pr(X = 2^k m \mid X = L) = p_k$$

for k = . . . -1, 0, 1, . . .; also, because $\{\ldots, p_{-1}, p_0, p_1, \ldots\}$ defines a probability distribution, $\sum_{k=-\infty}^{\infty} p_k = 1$. The 2^k factor reflects the fact that the next-smaller and next-larger amounts are, respectively, half and double the amounts in the box chosen.

Now suppose P observes $X = 2^k m$ for some k. Then

$$Pr(X = 2^k m \mid X = S) = Pr(X = 2^{k+1} m \mid X = L) = p_{k+1}.$$

Because P observes X,

$$Pr(X = 2^k m \mid X = L) + P(X = 2^k m \mid X = S) = p_k + p_{k+1} > 0. \tag{4}$$

It follows from Bayes' Theorem that

$$Pr(X = L \mid X = 2^k m) = \frac{Pr(X = 2^k m \mid X = L)Pr(X = L)}{Pr(L = 2^k m \mid X = L)Pr(X = L) + Pr(S = 2^k m \mid X = S)Pr(X = S)}$$

$$= \frac{p_k}{p_k + p_{k+1}}. \tag{5}$$

The <u>exchange condition</u> given by (3) now becomes

$$\frac{p_k}{p_k + p_{k+1}} < \frac{2}{3}, \text{ or } \qquad p_{k+1} > \frac{p_k}{2}. \tag{6}$$

Thus, whatever P observes, an exchange is profitable iff the probability that P has picked the S box (p_{k+1}) is more than 1/2 the probability that P has picked the L box (p_k).

We next give three discrete distributions where (6) is satisfied at least some of the time:

1. Exchange condition usually satisfied. $m = 1$; $p_k = 1/4$ for $k = 0, \ldots, 3$; $p_k = 0$ for all other k. Condition (6) is satisfied for $x = 1/2$, 1, 2, and 4 but not for $x = 8$ because $p_4 \not> p_3/2$; the former four values are the only ones that satisfy (4). Condition (6) holds with probability 7/8, because, for (6) to fail, $(X = L)$ and $(L = 8)$ must both occur.

2. Exchange condition usually not satisfied. $m = 1$; $p_k = 2^{-k-1}$ for $k = 0, 1, 2, \ldots$; $p_k = 0$ for all other k. This is a probability distribution because

$$\sum_{k=0}^{\infty} \Pr(X = 2^{-k-1} \mid X = L) = \frac{1}{2} + \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{1 - 1/2} = 1.$$

Condition (6) is satisfied for $x = 1/2$ because $p_0 > p_{-1}/2$ but not for $x = 1, 2, 4, \ldots$. It is easy to show that (6) holds with probability 1/4.

3. Exchange condition always satisfied. $m = 1$; $p_{-k} = p_k = \frac{1}{7}\left(\frac{2}{3}\right)^{k-1}$ for $k = 1, 2, \ldots$; $p_0 = \frac{1}{7}$. This is a probability distribution because

$$\sum_{k=-\infty}^{\infty} \Pr(X = 2^k \mid X = L) = \frac{2}{7}\sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^{k-1} + \frac{1}{7} = \frac{2}{7}\left[\frac{1}{1 - 2/3}\right] + \frac{1}{7} = 1.$$

Condition (6) is always satisfied:

$k \le -2$: $p_{k+1} = (3/2)p_k > p_k/2$

$k = -1$: $p_0 = 1/7 > p_{-1}/2 = 1/14$

$k = 0$: $p_1 = 1/7 > p_0/2 = 1/14$

$k \geq 1$: $p_{k+1} = (2/3)p_k > p_k/2$

Note that example 3 is a distribution that forms a plateau at $x = 1/2$, 1, and 2, falling off in each direction from this plateau as $x \to 0$ and $x \to \infty$. The rate of decrease as x increases is slow enough, however, that condition (6) is always satisfied.

We now show

Theorem 1. For any $\{\ldots, p_{-1}, p_0, p_1, \ldots\}$, exchange condition (6) must hold for at least one value of k.

Proof. Suppose, to the contrary, that

$p_{k+1} \leq p_k/2$, or $p_k \geq 2p_{k+1}$

for $k = \ldots -1, 0, 1, \ldots$. Choose a k such that $p_k > 0$. If $n < k$, then $p_n \geq 2^{k-n}p_k$. But this means that as $p_n \to \infty$, $n \to -\infty$, contradicting the assumption that

$$\sum_{k=-\infty}^{\infty} p_k = 1. \square$$

Thus, no matter what the discrete distribution is, condition (6) sometimes holds. As example 3 illustrated, for some discrete distributions, switching is always profitable.

4. Continuous Distributions

We next consider the case when X, L, and S are continuous random variables. Because

$\Pr(S \leq x) = \Pr(L \leq 2x)$,

the cumulative distribution functions of S and L are related by

$$F_S(x) = F_L(2x). \tag{7}$$

Differentiating (7) demonstrates that the probability density functions are related by

$$f_S(x) = 2f_L(2x).$$

Now suppose that P, after picking B1, observes X, where $x \le X \le x + dx$. The conditional probability that B1 is the L box, given this observation, is analogous to the probability given by (5) in the discrete case:

$$\Pr(X = L \mid x \le X \le x + dx) = \frac{f_L(x)dx}{f_L(x)dx + f_S(x)dx} = \frac{f_L(x)}{f_L(x) + 2f_L(2x)}.$$

The continuous analogue of (3) is, therefore,

$$\frac{f_L(x)}{f_L(x) + 2f_L(2x)} < \frac{2}{3},$$

which is equivalent to <u>exchange condition</u>

$$f_L(x) < 4f_L(2x). \tag{8}$$

In the continuous case, then, an exchange is profitable iff, once $x \le X \le x + dx$ is observed, the density of L at x is less than four times the density of L at 2x. In other words, the value of $f_L(\cdot)$ at 2x must be more than one-quarter its value at x--it cannot decrease too rapidly as x increases.

We next give three continous distributions where (8) is satisfied at least some of the time:

4. Uniform distribution. For simplicity, let $f_L(x) = 1, 0 \le x \le 1$. Condition (8) is satisfied iff $x \le 1/2$, which occurs with probability 3/4 because, for (8) to fail, $(X = L)$ and $(L > 1/2)$ must both occur.

5. Exponential distribution. $f_L(x) = e^{-x}, 0 \le x \le \infty$. Condition (8) is satisfied iff $e^{-x} < 4e^{-2x}$, which is equivalent to $x < \ln 4 \cong 1.39$. It is easy to verify that $\Pr(L \le \ln 4) = 3/4$ and $\Pr(S \le \ln 4) = 15/16$, so the exchange condition is satisfied with probability

$$\Pr(X \le \ln 4) = (1/2)(3/4) + (1/2)(15/16) = 27/32 \cong .844.$$

6. Exchange condition always satisfied.

$$f_L(x) = \begin{cases} \dfrac{1-\epsilon}{2-\epsilon} & \text{if } 0 \le x \le 1 \\ \dfrac{1-\epsilon}{2-\epsilon} x^{-2+\epsilon} & \text{if } x > 1, \end{cases}$$

where $0 < \epsilon < 1$. This is a probability distribution because

$$\int_0^\infty f_L dx = \left(\frac{1-\epsilon}{2-\epsilon}\right) + \left(\frac{1-\epsilon}{2-\epsilon}\right)\left[\frac{x^{-1+\epsilon}}{-1+\epsilon}\right]_1^\infty = \left(\frac{1-\epsilon}{2-\epsilon}\right)\left[1 + \frac{1}{1-\epsilon}\right] = 1.$$

Condition (8) is always satisfied:

$$0 \le x \le \frac{1}{2}: \quad f_L(2x) = f_L(x) = \frac{1-\epsilon}{2-\epsilon} \Rightarrow f_L(2x) > \frac{1}{4} f_L(x)$$

$$\frac{1}{2} < x \le 1: \quad f_L(2x) = \left(\frac{1-\epsilon}{2-\epsilon}\right)(2x)^{-2+\epsilon} \ge \left(\frac{1-\epsilon}{2-\epsilon}\right)\left(\frac{1}{2^{2-\epsilon}}\right) > \frac{1}{4}\left(\frac{1-\epsilon}{2-\epsilon}\right) = \frac{1}{4} f_L(x)$$

$$x > 1: \quad f_L(2x) = \left(\frac{1-\epsilon}{2-\epsilon}\right)(2x)^{-2+\epsilon} \ge \left(\frac{1-\epsilon}{2-\epsilon}\right)\left(\frac{x^{-2+\epsilon}}{2^{2-\epsilon}}\right) > \frac{1}{4} f_L(x).$$

Like example 3, for which the exchange condition is always satisfied in the discrete case, example 6 is a distribution that forms a plateau (for $0 \le x \le 1/2$); from this plateau it falls off, but not too rapidly, as $x \to \infty$.

We now show

Theorem 2. For any probability density function $f_L(x)$, exchange condition (8) cannot fail for all values of x.

Proof. Suppose, to the contrary, that

$$f_L(x/2) \ge 4f_L(x) \text{ for all } x \in (0, \infty). \tag{9}$$

Find an interval $[2^k, 2^{k+1}]$ such that

$$\int_{2^k}^{2^{k+1}} f_L(x)dx = K > 0. \tag{10}$$

Now (9) and (10) imply that

$$\int_{2^{k-1}}^{2^k} f_L(x)dx = \int_{2^k}^{2^{k+1}} f_L(y/2)(1/2)dy \ge 4\left(\frac{1}{2}\right)\int_{2^k}^{2^{k+1}} f_L(y)dy = 2K.$$

By induction,

$$\int_{2^{k-n-1}}^{2^{k-n}} f_L(x)dx \ge 2^n K$$

for $n = 0, 1, 2, \ldots$. It follows that

$$\int_0^{\infty} f_L(x)dx \geq \sum_{n=0}^{\infty} \int_{2^{k-n}}^{2^{k-n+1}} f_L(x)dx \geq \sum_{n=0}^{\infty} 2^n K = \infty,$$

so $f_L(x)$ cannot be the density function of a continuous random variable.□

Thus, no matter what the continuous distribution is, condition (8) sometimes holds. As example 6 illustrated, for some continuous distributions, exchange is always profitable.

5. When Is Envy Rational?

We have shown that, for both discrete and continuous probability distributions, it is always possible for P to find some amount in B1 that provides grounds for envy--it pays, on the average, to exchange this amount for either half or twice the amount in B2. More surprising, for some distributions envy is always justified: every amount in B1 is less than the expected value of the amount in B2, making a switch always profitable.

The latter phenomenon can occur, however, only if the probability distribution has unbounded support. That is, no matter what amount P finds in B1, there is some possibility that the amount in B2 will be larger.

While necessary, however, a distribution with unbounded support is not a sufficient condition for always switching. The distribution must also not tail off "too fast" as the amount x in a box approaches infinity. Thus in the continuous case, the density at 2x be at least one-quarter of its value at x, as indicated by (8). In the discrete case, as indicated by (6), an exchange is profitable when the probability of a value at x is at least one-half the preceding probability. For the four examples of distributions with unbounded support, examples 3 and 6 always satisfy these conditions, but examples 2 and 5 satisfy them only sometimes.

Our results on the box problem suggest how gambling may well have a rational basis in envy. If one estimates that the hand one is dealt--literally in cards but figuratively in life--will, by a switch, become half or twice as valuable, a switch may be worthwhile.

In material terms, of course, there is never the possibility of an unbounded payoff, so P cannot always benefit, as in examples 3 and 6, from a switch. In spiritual terms, however, such a payoff cannot be ruled out (Brams, 1983). Indeed, Pascal postulated as much in his famous wager (Pascal, 1670): if believing in God when He exists provides an infinite spiritual reward, one should believe, or at least keep an open mind. Similarly, in the box problem one should switch, whatever one observes, if (6) or (8) is always satisfied.

Even for finite distributions, like those in examples 1 and 4, switching may be rational most of the time. Envy , therefore, is frequently justified. But, curiously enough, this is not the case if there is another player who must also agree to an exchange in a two-person game in which the probability distribution is common knowledge (Brams, Kilgour, and Davis, 1990). Thus in example 4, if P1 chooses B1 and P2 chooses B2, an exchange is in equilibrium not when the amount in both boxes is less than 1/2, which will be true one-quarter of the time, but only when each contains 0, which is a zero-probability event. Hence, neither player will ever offer to switch, however small the amount in its box, in equilibrium.

In nongame situations, our results suggest a rational basis for envy. Nonetheless, we do not pose them as a model of human behavior--more detailed arguments in specific situations are needed to make such a case. Instead, like the St. Petersburg paradox, which challenges the equating of utility and money (in this paradox, a fair bet requires an infinite expected bankroll), the box problem clarifies thinking about gambling and risk, albeit in a different context. Specifically, it adds an object of envy (i.e., B2), with unknown contents, and indicates what beliefs, in the form of a probability distribution, will fuel this envy.

Whereas the resolution of the St. Peterburg paradox requires that a gambler's utility function for increasing returns be bounded (French, 1986), the box problem requires the assumption of an a priori probability distribution. But this latter assumption does not so much resolve the inconsistency of P's wanting to switch, whichever box is chosen. Rather, by postulating such a distribution, the conditions under which P will be envious are made perspicuous. What remains anomalous, perhaps, is that some probability distributions encourage envy most if not all the time.

The St. Petersburg paradox stimulated Bernoulli (1738) and later theorists to construct foundations on which to build a theory of utility and risk. Perhaps the box problem will stimulate the development of new foundations for viewing these issues in a different light.

REFERENCES

Bernoulli, Daniel (1738). "Exposition of a New Theory on the Measurement of Risk" (English translation). Econometrica 22 (1954): 23-36.

Brams, Steven J. (1983). Superior Beings: If They Exist, How Would We Know? Game-Theoretic Implications of Omniscience, Omnipotence, Immortality, and Incomprehensibility. New York: Springer Verlag.

Brams, Steven J., D. Marc Kilgour, and Morton D. Davis (1990). "Unraveling in Exchange Games." Preprint.

Pascal, Blaise (1670). Pensées (English translation by A. J. Krailsheimer). New York: Penguin, 1966.

French, Simon (1986). Decision Theory: An Introduction to the Mathematics of Rationality. New York: Halsted/Wiley, 1986.