International Journal of Game Theory (1993) 22: 89–95

On a Deception Game with Three Boxes

KING-TAK LEE

Department of Mathematics, University of Malaya, 59100 Kuala Lumpur, Malaysia

Abstract: This paper clarifies the status of a deception game posed by Spencer. We show the value of this game is zero.

1 Introduction

Consider the following two-person zero-sum game Γ_n where $n \geq 2$ is an integer. Players 1 and 2 are both informed that n numbers have been drawn, each uniformly and independently from the interval $[-1, 1]$. Only player 2 is further informed the actual outcome of the numbers. These numbers are then dropped into n numbered boxes, one in each box. Player 2 closes the boxes, and labels the lids. At least $n-1$ of the lid-labels must equal to the contained numbers, but one of the lids may be labelled arbitrarily from $[-1, 1]$. Player 1 then looks at the labels, chooses a box, and receives from player 2 an amount equal to the selected box's actual contents.

The interesting feature here concerns the usefulness of certain information given a piece of it has been manipulated for deception. Obviously, by ignoring the labels completely, player 1 can guarantee himself an expected payoff of 0 by choosing any box. Thus $v_n \geq 0$ where v_n is the value of Γ_n. Notice that if player 2 is not allowed to falsify any number, player 1 can get an expected payoff greater than 0 by choosing the box with the largest lid-label, thus obtaining the expectation of the maximum of the ordered statistics. So the true information in the boxes is valuable to player 1, and asking whether $v_n = 0$ is tantamount to asking whether the falsification of one out of n pieces of information is enough to render the information completely useless to him. It is easy to see $v_2 = 0$ since player 2 can make the two boxes show the same lid-label. We will show later that $v_n > 0$ for $n \geq 4$. So falsifying one out of two pieces of information can make the information useless. When there are four or more pieces of information, falsifying only one cannot make the information useless. What is then the situation for precisely three pieces of information? This is the problem addressed here and the answer is: In fact, falsifying one out of three pieces of information can make the whole information useless.

The deception game Γ_n is due to Mark Thompson. The question whether $v_3 = 0$ was discussed but not solved in his Harvard undergraduate thesis in 1970. Subsequently, Joel Spencer (1973) posed this question as an open problem in the *American Mathematical Monthly.* He also mentioned Daniel Kleitman and Shmuel Zamir

0020–7276/93/2/89–95 $ 2.50

had independently proved $v_4 > 0$. Baston and Bostock (1988) later filled an important theoretical gap by proving very general deception games always have a solution.

The original problem had the box contents drawn from [0, 1], but we have rescaled the interval here to $[-1, 1]$. This permits an easier description of an optimal strategies for player 2 in Γ_3. In fact, we shall present two optimal strategies for player 2. The elegant mixed optimal strategy in Section 2 is constructed based on an idea attributed to Thompson by Spencer. Thompson has apparently not published this result. Thus a major aim of the present paper is to clarify the status of this deception game in the open literature. We present the simplest pure optimal strategy we could find in Section 3, and we show $v_n > 0$ for $n \geq 4$ in Section 4.

2 A Mixed Optimal Strategy for Player 2

Let the ordered triples (x_1, x_2, x_3) and (y_1, y_2, y_3) denote the contents and lid-labels respectively of the three boxes. Let I denote the interval $[-1, 1]$. A pure strategy for player 2 is taken to be a Lebesgue measurable function from I^3 to I^3 which he uses to change (x_1, x_2, x_3) to (y_1, y_2, y_3) such that $y_k = x_k$ for at least two of the values $k = 1, 2, 3$. A mixed strategy for player 2 is a probability measure over the set of his pure strategies. To prove $v_3 = 0$, it suffices to construct a strategy for player 2, pure or mixed, that can hold the expected payoff of player 1 to 0. A sufficient condition on such a strategy is that, for each (y_1, y_2, y_3) in I^3, the conditional expectation $E[x_k | (y_1, y_2, y_3)]$ is independent of k, that is, when player 1 sees (y_1, y_2, y_3), he is indifferent between choosing any one of the three boxes. We will refer this later as the indifferent property.

Let a, b and c denote the absolute values of the box contents, ordered from the largest to the smallest. Let $p = (ab + bc)/(ab + ac)$ and $q = (ac + bc)/(ab + ac)$. Figure 1 describes a mixed optimal strategy in which player 2 simply uses the probabilities within the rectangle to change the sign of one of the box contents. The signs + and − denote positive and negative box contents; either sign may be used if the contents is 0. If $a = 0$ then necessarily $b = c = 0$. Notice p and q are not properly defined when $a = 0$ or $b = c = 0$. For the first case we take the row pattern $a, b, c = 0, 0, 0$ to be −, −, −. For the second case we may assume $a > 0$. We then take the row pattern $a, b, c = a, 0, 0$ to be +, −, − or −, +, − depending whether the "a" box contents is positive or negative, and define $p = 1$.

Box contents (a b c) \ Lid-labels (a b c)	− − −	+ − −	− + +	+ + +
− − −		1		
+ − −	1			
− + −	p		1-p	
− − +	q		1-q	
+ + −		1-q		q
+ − +		1-p		p
− + +				1
+ + +			1	

Fig. 1. A mixed optimal strategy for player 2.

The verification that this is an optimal strategy involves checking that, given an observation of any of the four column patterns, player 1 will receive the same expected payoff irrespective of which box he chooses. Thus the indifferent property mentioned earlier is satisfied.

3 A Pure Optimal Strategy for Player 2

Two mathematically-interesting questions remain concerning Γ_3. Since player 1 has a pure optimal strategy, does player 2 also have a pure optimal strategy? What does the full set of the optimal strategies of player 2 look like? We can answer the first question in the affirmative, but we have no answer to the second question. The second problem is equivalent to that of characterizing all the extreme optimal strategies for player 2. A pure optimal strategy is obviously an extreme optimal strategy, but the converse is not true. Even the restricted problem of determining all his pure optimal strategies lies beyond our reach.

To construct a pure optimal strategy f for player 2, we introduce the following notation. For x, y and z in I, let $[x, y, z]$ denote all the distinct permutations on x, y and z, written as ordered triples. We call $[x, y, z]$ a combination. Two combinations are equal if and only if they admit the same permutations; thus $[x, y, z] = [x, z, y]$ and so on. Let the box contents be x, y and z, not necessarily given according to the order of the boxes. If f changes z to t, we denote this symbolically by $[x, y, z] \to [x, y, t]$. We use the following procedure to describe f and verify it is an optimal strategy.

(i) Define a partition $\{P_k\}$ for I^3;
(ii) Define f on each P_k. Take a *fixed* combination, say $[x, y, z]$, in P_k. If $[x, y, z] \to [x, y, t]$ by the rule f, verify t is in I;
(iii) Obtain S, the set consisting of all distinct combinations in P_k which are mapped by f to $[x, y, t]$;
(iv) From each combination in S, extract the appropriate triple or triples. Verify these triples satisfy the indifferent property, that is, the sum of these triples is a triple whose components are all equal.

We will illustrate the above procedure for one particular case below, and leave the verification of the other cases to the reader. Let A_k denote the set of triples (or combinations) in I^3 with exactly k distinct components. Hence $I^3 = A_1 \cup A_2 \cup A_3$. We first deal with the more difficult case A_3 where all the triples have distinct components. Let

$$B_1 = \{[x, y, z]:\ -1 \le x < y \le 0,\ y < z \le 1\},$$
$$B_2 = \{[-x,\ -y,\ -z]:\ -1 \le x < y < 0,\ y < z \le 1\},$$
$$C_1 = \{[x, y, z]:\ -1 \le x < y \le 0,\ y < z \le 1 + 2y\},$$
$$C_2 = \{[x, y, z]:\ -1 \le x < y \le 0,\ 1 + 2y < z \le 1\}.$$

Then A_3 is partitioned by B_1 and B_2, and B_1 is partitioned by C_1 and C_2. We hereafter assume $x < y < z$.

For $[x, y, z] \in C_1$, change z to t where

$$t = \begin{cases} x & \text{if } x + z = 2y, \\ 2y - z & \text{if } x + z \ne 2y. \end{cases}$$

We now illustrate our procedure for the case $x + z \ne 2y$.

(ii) It is easy to verify $t = 2y - z \in I$;
(iii) $S = \{[x, y, z],\ [2y - z, y, 2y - x]\}$. By our assumption, $[x, y, z] \in S$. Since $-1 \le 2y - z < y \le 0$ and $y < 2y - x \le 1 + 2y$, $[2y - z, y, 2y - x] \in C_1$. Furthermore, $(2y - z) + (2y - x) \ne 2y$. By the definition of f on this combination, $2y - x$ is changed to $2y - (2y - x) = x$. The reader can verify the two combinations in S are distinct, and no other combinations in C_1 can be mapped by f to $[x, y, 2y - z]$;
(iv) Extract the triples (x, y, z) and $(2y - x, y, 2y - z)$. These two triples, which are mapped by f to $(x, y, 2y - z)$, satisfy the indifferent property. It is clear the remaining five pairs of triples also satisfy this property.

For the case $x + z = 2y$, $S = \{[x, y, z]\}$.

We may take $y < 0$ in C_2 since C_2 is empty if $y = 0$. Let $C_2 = C_2' \cup (C_2 - C_2')$ where

$$C_2' = \{[x, y, z]:\ -1 < x < y < 0,\ 1 + 2y < z < 1\}.$$

For $[x, y, z] \in C_2'$, change z to t where

$$t = \begin{cases} 1+x+y-z & \text{if } z-2x-2y \le 3, \\ -(1+z)/2 & \text{if } z-2x-2y > 3. \end{cases}$$

We list some properties of t which are required to verify f works.

$$-1 < t < 0;$$
$$1+2x < 2t+z-2y < 2t+z-2x < 1.$$

There are two separate cases here.

Case 1: $t \ne x$ and $t \ne y$. In this case $S = \{[x, y, z], [x, t, 2t+z-2y], [y, t, 2t+z-2x]\}$. Notice t is invariant when x and y are interchanged. This means we do not have to check, for example, whether x or t is larger in $[x, t, 2t+z-2y]$. We have written the combinations in S such that the last component is the one to be changed.

Case 2: $t = x$ or $t = y$. We summarize the results to be verified in Figure 2 where $-1 < x < y < 0$.

Conditions	Combinations in C_2'	Mapped by f to	Combinations required in A_2
$2x+y \ge -2$	$[x, y, 1+y]$	$[x, y, x]$	$[x, 1+2x-y, x]$
$2x+y < -2$	$[x, y, -1-2x]$	$[x, y, x]$	$[x, -1-2y, x]$
$x+2y \ge -2$ and $2y < x$	$[x, y, 1+x]$	$[x, y, y]$	$[1-x+2y, y, y]$
$x+2y < -2$	$[x, y, -1-2y]$	$[x, y, y]$	$[-1-2x, y, y]$

Fig. 2. The case $t = x$ or $t = y$.

We have to verify, for example, if $-1 < x < y < 0$ and $2x+y \ge -2$, then $[x, y, 1+x] \in C_2'$, and is the only combination in C_2' mapped by f to $[x, y, z]$. We omit these verifications because they are straightforward. We now explain why we include column 4 in Figure 2. Recall we define f on C_2' by $[x, y, z] \to [x, y, t]$. For certain values of x, y and z, we have $t = x$ or $t = y$. For the sake of arguments, suppose the first set of conditions holds so that $[x, y, 1+y]$ is mapped to $[x, y, x]$. To "balance" this combination, we require a combination in A_2, namely $[x, 1+2x-y, x]$, and map it also to $[x, y, x]$ so that a total of two combinations in A_2 and C_2' are mapped to $[x, y, x]$. These two combinations contain nine triples which are then put into three groups, each with three triples. We then verify each group of triples satisfies the indifferent property. Let D_1 denote the set of combinations in A_2 required to balance the combinations in C_2' for all possible values of x and y. It is important to check all the combinations in D_1 are distinct to ensure no combination in D_1 is required more than once.

The expected payoff received by player 1 is equal to a certain Lebesgue integral evaluated over I^3. This is in turn equal to the integral evaluated over I^3 minus any set with measure 0. For $[x, y, z] \in C_2' - C_2$, $x = -1$ or $z = 1$. It follows that $C_2 - C_2'$ is part of the boundary of I^3 and therefore has measure 0. Hence f may be defined arbitrarily on $C_2 - C_2'$. Put it another way, the event that the contents of the three boxes fall in $C_2 - C_2'$ has probability measure 0. This event may be disregarded without affecting the computation of the expected payoff to player 1.

For $[-x, -y, -z]$ in B_2, define $f[-x, -y, -z] = -f[x, y, z]$. That is, for $[-x, -y, -z]$ in B_2, change $[x, y, z]$ according to f defined on B_1 and then reverse the sign of each component. It is clear that, since f works for combinations in B_1, it also works for the combinations in B_2. In this case the set of combinations required in A_2 is D_2, given by $D_2 = -D_1$. It can be checked $D_1 \cap D_2$ is empty.

We have defined f on the subsets D_1 and D_2 of A_2. We now define f on $A_2 - D_1 - D_2$ by $[x, x, u] \to [x, x, x]$, and we leave all combinations unchanged in A_1. This completes the construction of a pure optimal strategy for player 2.

4 Four or More Boxes

We show $v_n > 0$ for $n \geq 4$, that is, the falsified information is useful to player 1 when there are four or more boxes. Toward this goal, we will construct a strategy for player 1, using only the lid-labels (y_1, y_2, y_3, y_4) of the first four boxes, which guarantees an expected payoff strictly greater than 0. The basic idea here is player 1 can exploit the situation when the contents (x_1, x_2, x_3, x_4) of these four boxes fall in the neighborhood of $(-1, -1, 1, 1)$ (or $(-1, 1, -1, 1)$ et cetera).

Let ε be a small positive number ($\varepsilon = 0.1$ will do). Define

$$
\begin{aligned}
G &= \{(x_1, x_2, x_3, x_4) \in I^4 : x_1 \leq -1+\varepsilon,\ x_2 \leq -1+\varepsilon,\ x_3 \geq 1-\varepsilon,\ x_4 \geq 1-\varepsilon\}, \\
H &= \{(x_1, x_2, x_3, x_4) \in I^4 : x_1 \leq -1+\varepsilon,\ x_2 \leq -1+\varepsilon,\ x_3 \geq 1-\varepsilon\} \\
&\cup \{(x_1, x_2, x_3, x_4) \in I^4 : x_1 \leq -1+\varepsilon,\ x_2 \leq -1+\varepsilon,\ x_4 \geq 1-\varepsilon\} \\
&\cup \{(x_1, x_2, x_3, x_4) \in I^4 : x_1 \leq -1+\varepsilon,\ x_3 \geq 1-\varepsilon,\ x_4 \geq 1-\varepsilon\} \\
&\cup \{(x_1, x_2, x_3, x_4) \in I^4 : x_2 \leq -1+\varepsilon,\ x_3 \geq 1-\varepsilon,\ x_4 \geq 1-\varepsilon\}.
\end{aligned}
$$

Let player 1 adopt the mixed strategy: Choose box k ($k = 1, 2, 3, 4$) with probability $(1+y_k)/(4+y_1+y_2+y_3+y_4)$ if $(y_1, y_2, y_3, y_4) \in H$; choose each of the four boxes equiprobably otherwise. We leave it to the interested reader to verify a best response from player 2 against this mixed strategy is:

(i) If $(x_1, x_2, x_3, x_4) \in G$, change the smaller of x_1 and x_2 to 1;
(ii) If $(x_1, x_2, x_3, x_4) \in H - G$, change the tuple to any feasible tuple in $I^4 - H$;
(iii) If $(x_1, x_2, x_3, x_4) \in I^4 - H$, leave the tuple unchanged.

In case (ii), a feasible tuple refers to an (y_1, y_2, y_3, y_4) satisfying $y_k = x_k$ for at least three of the values $k = 1, 2, 3, 4$. It can be checked that player 1 receives more than $(x_1+x_2+x_3+x_4)/4$ if $(x_1, x_2, x_3, x_4) \in G$, and receives $(x_1+x_2+x_3+x_4)/4$ if

$(x_1, x_2, x_3, x_4) \in I^4 - G$. Since G has a positive measure, the expected payoff to player 1 is strictly greater than 0. This imples $v_n > 0$ for $n \geq 4$.

Acknowledgement: The author would like to thank the referees for several helpful suggestions.

References

Baston VJ and Bostock FA (1988) Deception Games. International Journal of Game Theory 17: 129–134

Spencer J (1973) A Deception Game. American Mathematical Monthly 80: 416–417

Received February 1991
Revised version January 1992