MARK J. MACHINA

THE TWO ERRORS: A SUMMARY

Upon reading Professor Allais' extensive reply, I feel that the entire controversy can be summed up by the following two points:

- Each error in the proof of the "Allais Impossibility Theorem" (Allais (1988)) stems from taking an equation that *is* valid in one setting (and one set of assumptions), and invoking it in a different setting (with different assumptions) where the equation *is not* valid.
- During our 1983–1986 correspondence and conference discussions, I expressed my agreement with these specific equations, without ever checking to see if they would still be valid under the different setting and assumptions of Allais' proof.

These points account for the ultimate failure of our correspondence, and for Allais' justifiable impression that I was *agreeing* with the key equations of his proof yet *disagreeing* with its conclusion.[1] The equations in question are (3*) and (6*) below, which appear as (23) and (20) in Allais' proof (1988, pp. 377–382).[2] As shown below, checking the validity of these equations under Allais' own assumptions would have been a simple thing to do, so I regret not having done it at the time. All I can say is that I *would* have personally checked the validity of each equation under these different assumptions, if the theorem in question had been my own.

But mathematically speaking, an invalid theorem is still an invalid theorem, and I fully stand behind by paper "Two Errors in the 'Allais Impossibility Theorem' ", which I have not changed upon reading Allais' reply. For the benefit of the reader, I sum up by redescribing each error as an instance of "using the right equation under the wrong assumptions."

The dispute concerns preferences over the set of all finite-outcome probability distributions $\mathbf{P} = (x_1, ..., x_n, p_1, ..., p_n)$ on an outcome

Theory and Decision **38:** 301–307, 1995.

interval $[0, M]$, where unless stated otherwise, n ranges over all positive integers. In particular, it centers on the following concept.[3]

DEFINITION. Given a preference functional $V(\cdot)$ over all finite-outcome probability distributions, the **local utility function** of $V(\cdot)$ at a lottery $\mathbf{P} = (x_1, ..., x_n, p_1, ..., p_n)$ is given by

$$U(x; \mathbf{P}) \stackrel{def}{\equiv} \frac{\partial V(\mathbf{P})}{\partial\, \text{prob}(x)} \equiv \left. \frac{\partial V(x_1, ..., x_n, x, p_1, ..., p_n, p_x)}{\partial p_x} \right|_{p_x=0} \quad \text{for all } x \in [0, M]. \qquad (1^*)$$

Note that $U(x; \mathbf{P})$ is defined over *all* values of x in $[0, M]$, whether or not they lie in the support $\{x_1, ..., x_n\}$ of $\mathbf{P}$. When $x = x_i \in \{x_1, ..., x_n\}$, we can substitute this into (1*), invoke the identity

$$V(x_1, ..., x_i, ..., x_n, x_i, p_1, ..., p_i, ..., p_n, p_x) \equiv V(x_1, ..., x_i, ..., x_n, p_1, ..., p_i + p_x, ..., p_n) \qquad (2^*)$$

and (1*) reduces to the simple formula

$$U(x_i; \mathbf{P}) = \partial V(x_1, ..., x_n, p_1, ..., p_n)/\partial p_i.$$

1. THE FIRST ERROR: LOCAL UTILITY FUNCTIONS AREN'T ALWAYS AFFINELY EQUIVALENT

The first error in Allais' proof concerns conditions under which the following is, or is not, valid:

If the preference functional $V^*(\cdot)$ is derived from $V(\cdot)$, then their local utility functions $U^*(\cdot;\cdot)$ and $U(\cdot;\cdot)$ will be affinely equivalent, in the sense that at each $\mathbf{P}$

$$U^*(x; \mathbf{P}) \equiv \lambda \cdot U(x; \mathbf{P}) + \mu \quad \text{for all } x \in [0, M] \qquad (3^*)$$

where neither λ or μ depend upon x, although they could depend upon $\mathbf{P}$.

A Valid Setting for Equation (3*):
Equation (3*) is valid whenever $V^*(\cdot)$ is derived from $V(\cdot)$ *by a monotonic transformation of the preference functional* – that is,

if we have the identity $V^*(\mathbf{P}) \equiv f(V(\mathbf{P}))$ for some increasing (differentiable) function $f(\cdot)$. To see this, differentiate this identity to obtain

$$U^*(x;\mathbf{P}) \overset{def}{\equiv} \frac{\partial V^*(\mathbf{P})}{\partial\, \mathrm{prob}(x)} \equiv f'(V(\mathbf{P}))\cdot \frac{\partial V(\mathbf{P})}{\partial\, \mathrm{prob}(x)}$$

$$\text{(4*)} \qquad \equiv f'(V(\mathbf{P}))\cdot U(x;\mathbf{P})$$

which yields the affine invariance equation (3*), where $\lambda = \lambda(\mathbf{P}) = f'(V(\mathbf{P}))$.[4]

Allais' Invalid Setting for Equation (3*):

But Allais' proof *does not involve* a monotonic transformation of the preference functional. In his case, $V^*(\cdot)$ is derived from $V(\cdot)$ *by a monotonic transformation* $y = f(x)$ *of the outcome variable* (1988, p. 378, eq.(7)). So instead of the identity $V^*(\mathbf{P}) \equiv f(V(\mathbf{P}))$, we have the identity $V^*(y_1,\ldots,y_n,p_1,\ldots,p_n) \equiv V(f^{-1}(y_1),\ldots,f^{-1}(y_n), p_1,\ldots,p_n)$. Allais' proof rests on the assertion that in such a case my theory would still imply – or I say it would imply – the affine invariance identity (3*) *"for any value of x* $(0 \le x \le M)$" (1988, p. 381, eq.(23)). In his 1995 reply (Sect. 4.3) he repeats this point.

We can obtain the actual relationship between the local utility functions in this setting by differentiating the (new) identity linking $V^*(\cdot)$ and $V(\cdot)$. I have done this in equations (21)–(24) of "Two Errors ...". to obtain[5]

$$\text{(5*)} \qquad U^*(y;\mathbf{P}) \equiv \lambda\cdot U(f^{-1}(y);\hat{\mathbf{P}}) + \mu$$

for all values of y (both in and out of the support of **P**), where $\hat{\mathbf{P}} = (f^{-1}(y_1),\ldots,f^{-1}(y_n),p_1,\ldots,p_n)$ is the distribution of the variable x when $y = f(x)$ has distribution $\mathbf{P} = (y_1,\ldots,y_n,p_1,\ldots,p_n)$.[6] Thus, far from satisfying the affine relationship (3*) – where $U(\cdot;\cdot)$ and $U^*(\cdot;\cdot)$ are evaluated at the *same* outcome value (i.e., x) and at the *same* distribution (i.e., **P**) – the two local utility functions in the case Allais actually considers are linked via *nonlinear relationships* between the outcome values and the distributions at which they are evaluated. Accordingly, any assertion that my theory would yield the identity (3*) *in Allais' case of a monotonic transformation of the outcome variable* is incorrect, and the use of (3*) (his equation (23)) to lead up to the contradiction in the final part of Allais' proof (1988, p. 381) is invalid.[7]

2. THE SECOND ERROR: MOMENT REPRESENTATION FUNCTIONS CAN'T ALWAYS BE USED TO DERIVE LOCAL UTILITY FUNCTIONS

The second error in Allais' proof concerns conditions under which the following is, or is not, valid:

For fixed n, if $G(M_1, ..., M_{2n-1})$ is a (differentiable) moment representation function for a preference functional $V(\cdot)$, then the local utility function of $V(\cdot)$ will take the form

$$(6^*) \qquad U(x;\mathbf{P}) \equiv \sum_{k=1}^{2n-1} \frac{\partial G(M_1, ..., M_{2n-1})}{\partial M_k} \cdot x^k \quad \text{for all } x \in [0, M]$$

at every n-outcome distribution $\mathbf{P} = (x_1, ..., x_n, p_1, ..., p_n)$, where $M_k = M_k(\mathbf{P}) = \sum_{i=1}^{n} x_i^k \cdot p_i$.

A Valid Setting for Equation (6):*

Equation (6*) is valid whenever $G(M_1, ..., M_{2n-1})$ is *an exact, global representation of* $V(\cdot)$ *over its entire domain of finite-outcome distributions of all orders* – that is, whenever we have the identity

$$(7^*) \qquad \begin{aligned} V(\mathbf{P}) \equiv\ & G(M_1(\mathbf{P}), ..., M_{2n-1}(\mathbf{P})) \\ & \text{for all } \mathbf{P} = \{x_1, ..., x_m, p_1, ..., p_m) \\ & \text{and all } m = \{1, 2, ..., n, ...\} \end{aligned}$$

Equation (7*) describes the preferences of someone who evaluates *all* finite-outcome distributions solely on the basis of these $2n - 1$ moments. In fact, when (7*) holds, the local utility formula (6*) will be valid at *all* finite-outcome distributions $\mathbf{P}$, not just distributions with exactly n outcomes.[8] To see this, pick any finite-outcome $\mathbf{P} = (x_1, ..., x_m, p_1, ..., p_m)$ (with $m \gtreqless n$), and any outcome level x (in or out of the support of $\mathbf{P}$), and simply differentiate the identity (7*) with respect to prob(x), noting that $\partial M_k(\mathbf{P})/\partial\, \text{prob}(x) \equiv x^k$.[9]

Allais' Invalid Setting for Equation (6):*

But Allais' proof *does not involve* a function $G(M_1, ..., M_{2n-1})$ that exactly represents $V(\cdot)$ over finite-outcome distributions of all orders. In his case, the function $G(M_1, ..., M_{2n-1})$ is only assumed to be able to represent preferences *over the set of distributions with exactly n outcomes* (1988, p. 377, eq.(5)). In other words, instead of (7*) we have restricted identity

$$(8^*) \qquad \begin{aligned} V(\mathbf{P}) \equiv\ & G(M_1(\mathbf{P}), ..., M_{2n-1}(\mathbf{P})) \\ & \text{for all } \mathbf{P} = (x_1, ..., x_n, p_1, ..., p_n) \end{aligned}$$

Allais' proof rests on the assertion that in such a case my theory would still imply – or I say it would imply – the local utility formula (6*) over the entire interval $[0, M]$.[10]

But if $G(M_1, ..., M_{2n-1})$ does not represent $V(\cdot)$ over all finite-outcome distributions of all orders, it *cannot* be used to derive the local utility function $U(\cdot; \mathbf{P})$ over $[0, M]$, *even at an n-outcome distribution* $\mathbf{P}$. To see this, take any n-outcome distribution $\mathbf{P} = (x_1, ..., x_n, p_1, ..., p_n)$ and any outcome level x in $[0, M]$ but outside the support $\mathbf{P}$, so that by equation (1*) we have

$$(9^*) \quad U(x; \mathbf{P}) = \left. \frac{\partial V(x_1, ..., x_n, x, p_1, ..., p_n, p_x)}{\partial p_x} \right|_{p_x=0}$$

But the value of this right-hand derivative, even when evaluated at $p_x = 0$, depends upon the behavior of $V(\cdot)$ *over a set of* $(n+1)$*-outcome distributions*, namely distributions of the form $(x_1, ..., x_n, x, p_1, ..., p_n, p_x)$ for small positive p_x. Thus, if $G(M_1, ..., M_{2n-1})$ does not necessarily represent $V(\cdot)$ over distributions with more than n outcomes – as in Allais' setting – it cannot be used to evaluate the local utility function $U(\cdot; \mathbf{P})$ over $[0, M]$, even at an n-outcome distribution $\mathbf{P}$. Accordingly, any assertion that my theory would yield equation (6*) *under the conditions of Allais' theorem* is incorrect, and the use of (6*) (his equation (20)) to lead up to the contradiction in the final part of Allais' proof (1988, p. 381) is invalid.

Allais (1988, p. 401; Reply, Sect. 4.4) offers an alternative proof of his Impossibility Theorem – without a transformation $y = f(x)$, but instead based on applying the local utility formula (6*) to each of a *pair* of moment representation functions $G(M_q, ..., M_{q+2n-1})$ and $G^*(M_r, ..., M_{r+2n-1})$, with $r > q$. However, since each of these functions can still only represent preferences over n-outcome distributions, the argument of the previous paragraph continues to apply, and the alternative proof is similarly invalid.

In his Reply (Section 8.6), Allais further clarifies the setting and assumptions of his theorem:

"The particular case I have considered in my proof of the *Impossibility Theorem* is the one of a preference function *on which one has only two informations:*

a - the value of the preference function ϕ for the values of $x_1, x_2, \ldots, x_n; p_1, p_2, \ldots, p_n$ at the considered point (x_i, p_i) with $p_1 + p_2 + \cdots + p_n = 1$

b - the partial derivatives $\partial\phi/\partial p_i$ at the considered point (x_i, p_i)."

But this limited information about the function $\phi(\cdot)$ is *not enough to determine* its local utility function $U(\cdot; \mathbf{P})$ over all of [0, M], even at the considered distribution $\mathbf{P} = (x_1, \ldots, x_n, p_1, \ldots, p_n)$. *Any* function $\pi(\cdot; \mathbf{P})$ that satisfies $\pi(x_i; \mathbf{P}) = \partial\phi/\partial p_i$ (or $\pi(x_i; \mathbf{P}) = a \cdot \partial\phi/\partial p_i + b$) for $i = 1, \ldots, n$ is consistent with and implies the above information. While all such functions must be affinely equivalent on the set $\{x_1, \ldots, x_n\}$, they can completely differ over the rest of $[0, M]$. Allais' (1988) proof, which presents two such functions $\pi(\cdot; \mathbf{P})$ and $\pi^*(\cdot; \mathbf{P})$ and shows that they are not affinely equivalent *over the rest of* $[0, M]$, does not contradict any implication of my theory. By way of analogy: if we only know enough to assess *von Neumann–Morgenstern utility* at $x_1, \ldots, x_n$, it's no contradiction that this limited information is consistent with functions $U(\cdot)$ and $U^*(\cdot)$ that *differ over the rest of* $[0, M]$.

NOTES

[1] They also explain why Allais' reply objects to my having used the phrase "according to Allais" in describing equations (20) and (22) in my sketch of his 1988 proof. I am sorry that Professor Allais considers this wording inappropriate. In each case, "according to Allais" should be taken to read "Allais' published proof next presents and relies upon the following equation, which I have agreed with in our correspondence."

[2] A summary of the Allais (1988) proof, with cross-references to his original equation numbers, may be found in my "Two Errors ..." paper, and also in Allais' 1995 reply.

[3] "Two Errors ..." eqs. (7), (5) (first identity) and (27). This idea was first developed for preferences over the set of *all* univariate distributions in Machina (1982). My notation $V(\cdot)$ and $U(\cdot;\cdot)$ (from Machina (1982)) corresponds to Allais' $H(\cdot)$ and $\pi(\cdot;\cdot)$.

[4] This result appears in Machina (1988, Sect. 3). In Section 2, I showed that since any complete set of changes in the probabilities sums to 0, we can always add or subtract an arbitrary constant $\mu(\mathbf{P})$ to any local utility function.

[5] In those equations, the symbols λ and μ appeared as a and b. As before, they can depend upon $\mathbf{P}$ but not x.

[6] In the special case of expected utility, equation (5*) simply states the well-known fact that if $U(\cdot)$ is my von Neumann–Morgenstern utility function over the variable x, and $U^*(\cdot)$ is my utility function over the variable $y = f(x)$, then they satisfy the relationship $U^*(y) \equiv U(f^{-1}(y))$, or more generally, $U^*(y) \equiv \lambda \cdot U(f^{-1}(y)) + \mu$.

[7] The Appendix to Allais' 1995 reply contains a similar erroneous derivation of local utility in the case of a monotonic transformation of a *continuous* outcome variable: Allais' (A14) implies that his (A15) should read $\pi^{**}(y;\psi) = y + \mu$. This and his (correct) (A3) imply that $\pi^{**}(y;\psi) = y + \mu = e^{ln(y)} + \mu = \pi(ln(y);\varphi) + \mu$, so that (A16) should read "$\pi^{**}(y;\psi) = \pi(ln(y);\varphi) + \mu$, where $\varphi(\cdot)$ is the density of x when $y = f(x)$ has the density $\psi(\cdot)$."

[8] This has been shown by Peter Bardsley ("Local Utility Functions," *Theory and Decision*, March 1993) for the case of central moments, but his proof extends to the case of absolute moments as well.

[9] The example in Allais' 1995 Appendix actually involves an infinite-moment, density function analogue of this valid setting. There, he considers an *exact, global representation* (A5) in terms of an infinite number of moments, and finds that its moment-based local utility formula (A8) does indeed match the directly derived formula (A3). Recall however, that Allais does not intend his Impossibility Theorem to apply to such continuous distributions (1988, p. 401). Also, see Note 6 above for a subsequent error in this Appendix.

[10] Recall that formula (6*) appears as eq. (20) in Allais' proof (p. 380). The proof does not explicitly state whether it is to hold for $x \in \{x_1, ..., x_n\}$ or for $x \in [0, M]$. However, in the sketch of the proof on pp. 358–359, this same equation (there numbered as (42)) is stated to hold for $0 \leq x \leq M$. In addition, eq. (24) of his proof, which is stated to follow from eqs. (20), (22) and (23), holds "whatever the value of x over the whole range $(0, M)$" (p. 381). Inspection of the proof shows that (24) could not hold over this whole range unless (20) does also.

Department of Economics,
University of California, San Diego,
La Jolla, CA 92093-0508,
U.S.A.