This paper has appeared in Grazer Philosophische Studien 40 (1991), pp. 185 - 194.

SIMPSON'S PARADOX AND THE FISHER–NEWCOMB PROBLEM

Carl G. Wagner
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996-1300, U.S.A.

ABSTRACT. I argue that the Fisher smoking problem and Newcomb's problem are decision-theoretically identical, each having at its core an identical case of Simpson's paradox.

1. Introduction

The present essay [1] was stimulated by the paper "An Epistemic Principle Which Solves Newcomb's Paradox," by Keith Lehrer [2] and Vann McGee, and I am honored to have it appear along with their paper in this volume.

I argue here that the Fisher smoking problem and Newcomb's problem are decision-theoretically identical, each having at its core an identical case of Simpson's paradox for certain empirical probabilities (observed relative frequencies). From this perspective, the incorrect solutions to these problems arise from (1) treating the problem as cases of decisionmaking under risk, and (2) adopting certain "global" empirical conditional probabilities as one's subjective probabilities.

The correct solutions to these problems are based on either (1) treating them as cases of decisionmaking under uncertainty with lottery acts, and adopting

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX certain "local" empirical conditional probabilities as one's subjective probabilities (the approach that I consider to be the most natural), or (2) retaining the methodology of decisionmaking under risk, constructing one's subjective probabilities as certain weighted averages of the aforementioned local empirical conditional probabilities (the approached favored by Lehrer and McGee).

2. Fisher's Problem

Imagine that you are trying to decide whether to smoke (s) or not ($\overline{s}$). Bearing on your decision is the issue of whether you will get cancer (c) or not ($\overline{c}$), and whether you will experience nicotine stimulation (n) or not ($\overline{n}$). Assume that n if and only if s. In many experimental studies it has been shown that the set of subjects contracting cancer (C) is overrepresented among the set of smokers (S), i.e., that

(2.1) $$\mathcal{P}(C/S) > \mathcal{P}(C/\overline{S}).$$

(We follow here the convention that A denotes the set of experimental subjects of which the proposition a is true, and that $\mathcal{P}(A/B) = \text{card}(A \cap B)/\text{card}(B)$, the relative frequency of A in B, or conditional empirical probability of A, given B.) On the other hand, recent studies, while not challenging the values of $\mathcal{P}(C/S)$ and $\mathcal{P}(C/\overline{S})$, have controlled for the presence in subjects of a particular gene (g) or its absence ($\overline{g}$), and these studies reveal that

(2.2) $$\mathcal{P}(C/S \cap G) = \mathcal{P}(C/\overline{S} \cap G)$$

and

(2.3) $$\mathcal{P}(C/S \cap \overline{G}) = \mathcal{P}(C/\overline{S} \cap \overline{G})$$

(It follows that the common value of the relative frequencies appearing in (2.2) is $\mathcal{P}(C/G)$, and the common value of the relative frequencies appearing in (2.3) is $\mathcal{P}(C/\overline{G})$.)

The joint occurence of (2.1), (2.2), and (2.3) is an example of *Simpson's Paradox* [3] . While S is "globally" positively relevant to C for the empirical measure $\mathcal{P}$, it is "locally" independent of C (in more standard terminology, conditionally independent of C, given G and given $\overline{G}$). This state of affairs necessitates the non-independence of G and C, as well as the non-independence of G and S [4] , and, indeed, the experimental data reveal that $\mathcal{P}(C/G) > \mathcal{P}(C/\overline{G})$ and $\mathcal{P}(S/G) > \mathcal{P}(S/\overline{G})$. Thus the positive relevance of G to both C and S is what accounts for (2.1) [5] . In short, possession of the gene in question apparently predisposes one both to smoke and to get cancer. But, given that one has the gene, smoking does not increase the chance of getting cancer, and the same is true if one does not have the gene.

Assume that you do not know whether you possess the gene in question, and that your utility function u over the possible outcomes $c \wedge n$, $c \wedge \overline{n}$, $\overline{c} \wedge n$, and $\overline{c} \wedge \overline{n}$ satisfies the inequalities

$$(2.4) \qquad u(c \wedge n) > u(c \wedge \overline{n}) \qquad \text{and} \qquad u(\overline{c} \wedge n) > u(\overline{c} \wedge \overline{n}),$$

that is, other things being equal, you prefer nicotine stimulation to its absence. All of the foregoing considered, it seems unexceptionable for you to choose between smoking and not smoking within the analytical framework of decisionmaking under uncertainty with lottery acts, a generalization of classical decisionmaking under uncertainty in which acts map states of the world to subjective probability distributions over outcomes, rather than to outcomes *tout court* [6] . Here, of course, the acts are s and $\overline{s}$, the states of the world g and $\overline{g}$[7], and the outcomes are $c \wedge n$, $c \wedge \overline{n}$, $\overline{c} \wedge n$, and $\overline{c} \wedge \overline{n}$. Denote your four relevant subjective probability distributions by $p_s^{(g)}$, $p_{\overline{s}}^{(g)}$, $p_s^{(\overline{g})}$, and $p_{\overline{s}}^{(\overline{g})}$, where, for example, $p_s^{(g)}$ is your subjective probability distribution over outcomes, given that you have the gene and that you smoke [8] . In view of the scientific evidence and the assumption that n if and only if s, it seems unexceptionable for you to adopt the following as values of these distributions:

$$(2.5)\qquad \begin{array}{lcccc} & c\wedge n & c\wedge\overline{n} & \overline{c}\wedge n & \overline{c}\wedge\overline{n} \\ p_s^{(g)}: & \mathcal{P}(C/S\cap G) & 0 & \mathcal{P}(\overline{C}/S\cap G) & 0 \\ p_{\overline{s}}^{(g)}: & 0 & \mathcal{P}(C/\overline{S}\cap G) & 0 & \mathcal{P}(\overline{C}/\overline{S}\cap G) \\ p_s^{(\overline{g})}: & \mathcal{P}(C/S\cap\overline{G}) & 0 & \mathcal{P}(\overline{C}/S\cap\overline{G}) & 0 \\ p_{\overline{s}}^{(\overline{g})}: & 0 & \mathcal{P}(C/\overline{S}\cap\overline{G}) & 0 & \mathcal{P}(\overline{C}/\overline{S}\cap\overline{G}). \end{array}$$

You proceed to calculate the expected utility of each act under each possible state of the world, here denoted $E^{(g)}(s)$, $E^{(g)}(\overline{s})$, $E^{(\overline{g})}(s)$, and $E^{(\overline{g})}(\overline{s})$, where, for example, $E^{(g)}(s) = \mathcal{P}(C/S\cap G)u(c\wedge n) + \mathcal{P}(\overline{C}/S\cap G)u(\overline{c}\wedge n)$. To complete this exercise according to the standard procedure for solving a problem of decisionmaking under uncertainty with lottery acts, you must determine your subjective probability distribution q over the possible states of the world g and $\overline{g}$, and then choose between s and $\overline{s}$ based on the magnitudes of the weighted averages

$$(2.6)\qquad \mathcal{E}(s) := q(g)E^{(g)}(s) + q(\overline{g})E^{(\overline{g})}(s)$$

and

$$(2.7)\qquad \mathcal{E}(\overline{s}) := q(g)E^{(g)}(\overline{s}) + q(\overline{g})E^{(\overline{g})}(\overline{s})$$

In the present case, however, it follows from (2.2), (2.3), (2.4) and (2.5) that

$$(2.8)\qquad E^{(g)}(s) > E^{(g)}(\overline{s}) \quad\text{and}\quad E^{(\overline{g})}(s) > E^{(\overline{g})}(\overline{s}),$$

so $\mathcal{E}(s) > \mathcal{E}(\overline{s})$ for *every* probability distribution q over g and $\overline{g}$, mooting the assessment of q. (This is the dominance principle for decisionmaking under uncertainty with lottery acts). It follows that the rational course of action is for you to smoke.

But wait. The utility inequalities in (2.4) are not the only ones that hold for you. Surely, other things being equal, you prefer no cancer to cancer, i.e.,

$$u(\overline{c} \wedge n) > u(c \wedge n) \quad \text{and} \quad u(\overline{c} \wedge \overline{n}) > u(c \wedge \overline{n}). \tag{2.9}$$

Let us suppose, moreover, that u also satisfies the inequality

$$\begin{aligned} \mathcal{P}(C/S)u(c \wedge n) + \mathcal{P}(\overline{C}/S)u(\overline{c} \wedge n) \\ < \mathcal{P}(C/\overline{S})u(c \wedge \overline{n}) + \mathcal{P}(\overline{C}/\overline{S})u(\overline{c} \wedge \overline{n}), \end{aligned} \tag{2.10}$$ [9]

the interpretation of which will be clarified shortly.

Now you are also acquainted with the procedures involved in decisionmaking under risk, where each act amounts to a choice of a probability distribution over outcomes. You are curious to see how things come out under this apparently simpler procedure. In this case you need to assess subjective probability distributions p_s and $p_{\overline{s}}$ over the outcomes $c \wedge n$, $c \wedge \overline{n}$, $\overline{c} \wedge n$, and $\overline{c} \wedge \overline{n}$. Without much reflection you adopt the following as values of these distributions:

$$\begin{array}{lccccc} & & c \wedge n & c \wedge \overline{n} & \overline{c} \wedge n & \overline{c} \wedge \overline{n} \\ (2.11) & p_s: & \mathcal{P}(C/S) & 0 & \mathcal{P}(\overline{C}/S) & 0 \\ & p_{\overline{s}}: & 0 & \mathcal{P}(C/\overline{S}) & 0 & \mathcal{P}(\overline{C}/\overline{S}). \end{array}$$

Following the standard procedure for decisionmaking under risk, you then calculate the expected utilities of the acts s and $\overline{s}$, denoted $E(s)$ and $E(\overline{s})$, where, for example, $E(s) = \mathcal{P}(C/S)u(c \wedge n) + \mathcal{P}(\overline{C}/S)u(\overline{c} \wedge n)$. But it now follows from (2.10) that $E(s) < E(\overline{s})$, from which you seemingly must conclude that the rational course of action is for you *not* to smoke!

Choosing between the incompatible analyses presented above is not, however, a daunting task. In the first place, the treatment of this problem as a case of decisionmaking under risk is immediately supect, given the postulated uncertainty regarding your possession of the gene in question, and the results of controlling for G and $\overline{G}$ that revealed (2.2) and (2.3). Moreover, (2.2) and (2.3), along with the observations that $\mathcal{P}(C/G) > \mathcal{P}(C/\overline{G})$ and $\mathcal{P}(S/G) > \mathcal{P}(S/\overline{G})$, undermine the subjective probability assignments of (2.11). Indeed, one does not have, strictly speaking, a *single* pair $(p_s, p_{\overline{s}})$ of the subjective probability distributions over outcomes, but rather the double pair $\left(p_s^{(g)}, p_{\overline{s}}^{(g)}\right)$ and $\left(p_s^{(\overline{g})}, p_{\overline{s}}^{(\overline{g})}\right)$, the appropriate pair being determined by which state of the world, g or $\overline{g}$, obtains for you.

On the other hand, as Lehrer and McGee have argued (though with different notation), the single pair of subjective probability distributions $(p_s^*, p_{\overline{s}}^*)$, constructed as the weighted averages

$$p_s^* = q(g)p_s^{(g)} + q(\overline{g})p_s^{(\overline{g})} \tag{2.12}$$

and

$$p_{\overline{s}}^* = q(g)p_{\overline{s}}^{(g)} + q(\overline{g})p_{\overline{s}}^{(\overline{g})}, \tag{2.13}$$

(where $p_s^{(g)}$, etc., are given by (2.5), and q is your subjective distribution over the possibilities g and $\overline{g}$) can ground on approach based, at least formally, on the methodology of decisionmaking under risk. Indeed, it is easy to show (with $E^*(s)$ and $E^*(\overline{s})$ defined analogously to $E(s)$ and $E(\overline{s})$) that

$$E^*(s) = \mathcal{E}(s) \quad \text{and} \quad E^*(\overline{s}) = \mathcal{E}(\overline{s}) \tag{2.14}$$

where $\mathcal{E}(s)$ and $\mathcal{E}(\overline{s})$ are given by (2.6) and (2.7).

Since, as previously established, $\mathcal{E}(s) > \mathcal{E}(\overline{s})$ for every q, it follows that $E^*(s) > E^*(\overline{s})$ for every q, again mooting the assessment of q.

3. Newcomb's Problem

Newcomb's problem is now so well known among philosophers that it requires no elaborate exposition. A terse and thoroughly adequate description of this problem appears in any case in the paper of Lehrer and McGee in this volume. Those who wish further details will find them in Ells (1984).

In order to best exhibit the identity of form betwen the Fisher and Newcomb problems, I shall use the same symbols employed in the preceding section, but here interpreted as follows:

$s =$ you take both boxes

$\overline{s} =$ you take only the opaque box

$g =$ the predictor predicts that you'll take both boxes

$\overline{g} =$ the predictor predicts that you'll take only the opaque box

$n =$ you get the thousand dollars

$\overline{n} =$ you don't get the thousand dollars

$c =$ you don't get the million dollars

$\overline{c} =$ you get the million dollars

As in the preceding section, capital letters stand for sets of experimental subjects (here previous players of the game) with the obvious relation to the proposition denoted by the corresponding lower case letter. For example, S is the set of all previous players who took two boxes. Also, $\mathcal{P}(A/B)$ continues to denote the fraction card $(A \cap B)/$ card (B).

Finally, u denotes your utility function over the four possible outcomes $c \wedge n$, $c \wedge \overline{n}$, $\overline{c} \wedge n$, and $\overline{c} \wedge \overline{n}$.

If, as I claim, Newcomb's problem is identical with Fisher's problem, then it should be the case for Newcomb's problem that the background empirical probabilities display the features of Simpson's paradox in the form

(3.1.a) $$\mathcal{P}(C/S) > \mathcal{P}(C/\overline{S})$$

(3.1.b) $$\mathcal{P}(C/S \cap G) = \mathcal{P}(C/\overline{S} \cap G),$$

and

$$\text{(3.1.c)} \qquad \mathcal{P}(C/S \cap \overline{G}) = \mathcal{P}(C/\overline{S} \cap \overline{G}),$$

with the joint occurrence of (3.1.a)–(3.1.c) accounted for by the inequalities

$$\text{(3.2)} \qquad \mathcal{P}(S/G) > \mathcal{P}(S/\overline{G}),$$

and

$$\text{(3.3)} \qquad \mathcal{P}(C/G) > \mathcal{P}(C/\overline{G}).$$

In addition, u should satisfy

$$\text{(3.4)} \qquad u(c \wedge n) > u(c \wedge \overline{n}) \quad \text{and} \quad u(\overline{c} \wedge n) > u(\overline{c} \wedge \overline{n}),$$

$$\text{(3.5)} \qquad u(\overline{c} \wedge n) > u(c \wedge n) \quad \text{and} \quad u(\overline{c} \wedge \overline{n}) > u(c \wedge \overline{n}),$$

and

$$\text{(3.6)} \qquad \mathcal{P}(C/S)u(c \wedge n) + \mathcal{P}(\overline{C}/S)u(\overline{c} \wedge n) < \mathcal{P}(C/\overline{S})u(c \wedge \overline{n}) + \mathcal{P}(\overline{C}/\overline{S})u(\overline{c} \wedge \overline{n}).$$

Are (3.1)–(3.6) the usual assumed background of Newcomb's problem? There is clearly no problem with (3.4) and (3.5), but in the usual formulation (at least as extended by Eells (1984)), one has, in addition to (3.4) and (3.5), only the assumptions

$$\text{(3.7)} \qquad \mathcal{P}(G/S) > \mathcal{P}(G/\overline{S}),$$

and

$$\text{(3.8)} \qquad \mathcal{P}(G/S)u(c \wedge n) + \mathcal{P}(\overline{G}/S)u(\overline{c} \wedge n) < \mathcal{P}(G/\overline{S})u(c \wedge \overline{n}) + \mathcal{P}(\overline{G}/\overline{S})u(\overline{c} \wedge \overline{n}),$$

along with the stipulation that the predictor puts a million dollars in the opaque box if and only if he predicts that a person will take only that box. But this implies that $G = C$, which, along with (3.7) and (3.8), imply (3.1.a) and (3.6). Furthermore, (3.2) is equivalent to (3.7), and (3.3) holds in the form $1 > 0$ since $G = C$. The identity $G = C$ also clearly implies that $\mathcal{P}(C/S \cap G) = \mathcal{P}(C/\overline{S} \cap G) = 1$ and $\mathcal{P}(C/S \cap \overline{G}) = \mathcal{P}(C/\overline{S} \cap \overline{G}) = 0$, thus giving us (3.1.b) and (3.1.c) as well. It is because the foregoing conditional probabilities take the extreme values one and zero in the usual formulation of Newcomb's problem that the essential identity of the Newcomb and Fisher problems is obscured. In particular, the occurrence of Simpson's paradox at the core of Newcomb's problem is easy to miss.

The bonus from clarifying all this is that one realizes that the stipulation that the predictor puts a million dollars in the opaque box if and only if he predicts that person takes just that box is an inessential feature of Newcomb's problem, as unnecessary as stipulating in Fisher's problem that a person gets cancer if and only if he has the defective gene. So the predictor can use one of two chance devices to decide whether to leave the opaque box empty, one for players he predicts will take two boxes, and the other for players he predicts will take one box, so long as the probability of leaving the box empty is larger in the former case than in the latter.

Notes

1. Research supported in part by the National Science Foundation (DIR–8921269). This essay was written while the author was a Visiting Scholar in the Philosophy Department of the University of Arizona.
2. Though a mathematician by training and trade, I remained largely unaware of the most interesting foundational issues in probability and

decision theory until I was introduced to them by Keith Lehrer during the 1977 N.E.H. Summer Institute on Freedom and Causality at the Center for Advanced Study in the Behavioral Sciences. In the years since that fortunate meeting, Lehrer's influence on my intellectual and personal life have been profound and salutary, and I am delighted to be able to express my gratitude to him.

3. In general, Simpson's paradox is said to occur whenever $(A, \overline{A})$, $(B, \overline{B})$ and $(C_1, C_2, \ldots, C_n)$ are partitions of a sample X and $\mathcal{P}(A/B)\gamma\mathcal{P}(A/\overline{B})$, while, at the same time $\mathcal{P}(A/B \cap C_i)\lambda\mathcal{P}(A/\overline{B} \cap C_i)$ for $i = 1, \ldots, n$, where γ and λ are incompatible relations belonging to $\{\leq, <, =, \geq, >\}$. Real world cases of this phenomenon are so familiar in data analysis as to hardly retain a paradoxical air. See Simpson (1951), Bickel, *et al* (1975), and Wagner (1982).
4. For, as is easily shown (2.2), (2.3), and $\mathcal{P}(C/G) = \mathcal{P}(C/\overline{G})$ jointly imply that $\mathcal{P}(C/S) = \mathcal{P}(C/\overline{S})$. Similarly, (2.2), (2.3), and $\mathcal{P}(S/G) = \mathcal{P}(S/\overline{G})$ jointly imply that $\mathcal{P}(C/S) = \mathcal{P}(C/\overline{S})$.
5. Of course, positive relevance is not in general transitive.
6. See Fishburn (1981, 1988) and Anscombe and Aumann (1963).
7. By definition, states of the world must be such as to be uninfluenced by acts. This is clearly the case here, the matter of whether you possess the gene in question having been settled long before the chosen act, whichever it turns out to be, will take place.
8. Following the standard practice of economists (though not of philosophers) I regard the probability distributions $p_s^{(g)}$, etc., as being defined over outcomes *simpliciter*. I do *not* assume that there is some grand probability P defined over the boolean algebra generated by all acts, outcomes, and states of the world and take $p_s^{(g)}(o) = P(o/s \wedge g)$. Indeed,

I believe that one causes needless confusion by this approach, including puzzles about how to probabilize acts that are presumably within one's power to render true.

9. Given that $\mathcal{P}(C/S) > \mathcal{P}(C/\overline{S})$, there always exists a utility function u satisfying (2.4), (2.9) and (2.10). Here is one way to construct such a utility function. First choose arbitrary values for $u(c \wedge \overline{n})$ and $u(c \wedge n)$, with $u(c \wedge \overline{n}) < u(c \wedge n)$. If, for arbitrary $\alpha > 0$, we set $u(\overline{c} \wedge \overline{n}) = u(c \wedge \overline{n}) + \alpha$ and $u(\overline{c} \wedge n) = u(c \wedge n) + \alpha$, then u clearly satisfies (2.4) and (2.9). But by taking α sufficiently large, (2.10), which is easily seen to be equivalent to $u(c \wedge n) - u(c \wedge \overline{n}) < \alpha(\mathcal{P}(C/S) - \mathcal{P}(\overline{C}/S))$, can also be satisfied.

References

1. Anscombe, F. and Aumann, R. (1963), A definition of subjective probability, *Annals of Mathematical Statistics* **34**, 199–205.
2. Bickel, B., Hammel, E., and O'Connell, J. (1975), Sex bias in graduate admissions: data from Berkeley, *Science* **187**, 398–404.
3. Eells, E. (1984), Newcomb's many solutions, *Theory and Decision* **16**, 59–105.
4. Fishburn, P. (1981), Subjective expected utility: a review of normative theories, *Theory and Decision* **13**, 139–199.
5. Fishburn, P. (1988), *Nonlinear Preference and Utility Theory*, The Johns Hopkins University Press, Baltimore.
6. Lehrer, K. and McGee, V. (1990), An epistemic principle which solves Newcomb's paradox, *Research Report*, Department of Philosophy, University of Arizona, Tucson, AZ 85721.
7. Simpson, E. (1951), The intrepretation of interaction in contingency

tables, *Journal of the Royal Statistical Society, Series* B **13**, 238–241.

8. Wagner, C.H. (1982), Simpson's paradox in real life, *American Statistician* **36**, 46–47.