Simpson's Paradox and Causality[1]

Prasanta S. Bandyopadhyay, Mark Greenwood,
Don Wallace F. Dcruz, and Venkata Raghavan R.

[final draft]

Abstract

There are three types of questions associated with Simpson's Paradox (SP): (i) why is SP paradoxical? (ii) what conditions generate it? and (iii) what should be done about SP? Pertaining to the first two questions, we argue that SP has nothing to do with causality. However, causality plays a role in addressing the third question. Our research shows that one needs to divorce the question of the paradox itself and the reason it seems paradoxical from the question of what to do about it. By providing a logic-based account for the paradox, we critique and produce a counterexample to Spirtes, Glymour and Scheines' causal account of SP. We compare their approach to ours by means of two sets of experiments that show SP is not causal (Word count 129.)

Overview

[1]Several versions of the paper have been accepted for presentation in several places including in Germany, Holland, India, and the United States. This version of the paper was presented at the APA, Eastern Divisional Meetings and Jadavpur University. We would like to thank the audiences in those places including John G. Bennett, Gordon Brittan, Jr., Dan Flory, and James Mattingly, for their helpful comments.

Simpson's Paradox and Causality

> "The skeptic about causality pushes the brake pedal to make his car slow, flips a switch to make a lamp glow, puts his money in the bank to collect interest (Spirtes, Glymour, and Scheines, 2000, p.2.)"

> "[Physicists] continued to write equations in the office and talk cause-effect in the cafeteria…. [P]hysicists talk, write, and think one way and formulate physics in another (Pearl, 2009, pp.407-8.)"

Overview

Simpson's Paradox (SP) involves the reversal of the direction of a comparison or the cessation of an association when data from several sets are pooled. SP has wide applications in numerous disciples. Behind its applications and usefulness, several deeper issues have yet to be properly distinguished. Moreover, resolving one does not necessarily lead to the resolution of the rest. We will further argue that a conflation of those issues is in fact a factor in misreading the entire story about the paradox. Lately, however, it is almost conventional wisdom among scholars in this field to take the core of SP to be exclusively causal. In the words of Peter Spirtes, Clark Glymour, and Richard Scheines, "[t]he question is what *causal dependencies* can produce such a [case], and that question is properly known as "Simpson's paradox". (Spirtes, Glymour, and Scheines, 2000, p.40, emphasis is ours.)" Judea Pearl, for example, writes that "…the spice of Simpson paradox has turned out to be nonstatistical (i.e., causal) after all (Pearl, 2009, p.177.)" In a very recent paper, epidemiologists, Miguel Hernan, David Clayton, and Niels Keiding echo the same refrain:

"[Simpson] paradox and error arise only when the problem is stripped of its causal context and analyzed merely in statistical terms, or when non-causal concepts like … collapsibility [is] allowed to guide the analysis. Once the casual goal is made explicit and causal considerations are incorporated into the analysis, the course of action becomes crystal clear. (Hernan, Clayton, and Keiding in 2011, p. 784.)"

One purpose of this paper is to contest the conventional wisdom that traces SP to causality insofar as its central themes are concerned. Three questions need to be distinguished with regard to the paradox: (i) why or in what sense, is SP a paradox?, (ii) what are the conditions for the emergence of this paradox?, and (iii) what should one do when confronted with a typical case of the paradox (to be called hereafter the "what-to-do" question?). We will argue that SP has to do with causality only if we ask the "what-to-do" question. For the sake of brevity, we will confine ourselves to the views of Spirtes, Glymour, and Scheines' causal account, often called the CMU theorists' account, as our rejoinder, if it is correct, is adequately general to be applicable to any other causal accounts of the paradox.[2]

The first section of the paper contains examples of Simpson's paradox. In the second section, we will provide a logic-based account that addresses the first two questions about the latter. In the next section, we discuss Spirtes, Glymour, and Scheines' causal account of the paradox. In section four, we provide a counterexample to the causal account. Section five is devoted to a comparison between our account and the causal account.

[2]This does not, however, imply that there are no differences between the different casual accounts regarding Simpson's paradox. Pearl, for example, has developed a calculus of causality to handle causal cases including Simpson's paradox. In addition, according to both him and some other causal theorists, although the collapsibility principle goes hand in hand with the paradox it is not the central idea in unlocking its riddle, as the principle is fundamentally non-causal. To contrast Pearl's account with the CMU theorists', the CMU theorists have proposed a constraint on observational data so that the data do not generate Simpson's paradox, whereas Pearl does not offer such a constraint. Consequently, the need for the collapsibility principle does not arise for their account. For more on the CMU theorists' view, see, section 3. However, for our present purpose, what matters is the common assumption shared by both Pearl and the CMU theorists about the causal resolution of the paradox.

Here, we describe two experiments regarding Simpson's paradox and discuss their bearing on choosing between these two accounts. Section six evaluates an objection to our account. We conclude that the causal account fails to appreciate the significance of the three questions we outline above regarding the paradox.

1. Simpson's paradox:

Consider the following two examples of the paradox.

Simpson's Paradox (Type I)

Two Groups	Dept. 1		Dept. 2		Acceptance Rates		Overall Acceptance Rates
	Accept	Reject	Accept	Reject	Dept. 1	Dept. 2	
F	180	20	100	200	90%	33%	56%
M	480	120	10	90	80%	10%	70%

Table 1

Simpson's Paradox (Type II)

Two Groups	Dept. 1		Dept. 2		Acceptance Rates		Overall Acceptance Rates
	Accept	Reject	Accept	Reject	Dept. 1	Dept. 2	
Females	90	1410	110	390	6%	22%	10%
Males	20	980	380	2620	2%	12%	10%

Table 2

Table 1 represents an example of a formulation of the paradox in which the association in the subpopulations (departments) with higher acceptance rate for females is *reversed* in the combined population, with overall higher rates for males. Table 2 contains an example of an apparently paradoxical effect when the association between "gender" and "acceptance rates" in the subpopulations *ceases* to exist in the combined population.

Although the acceptance rates for females are higher in each department, in the combined population, those rates cease to be different.

2. Our logic-based account of the paradox:[3]

We begin with an analysis of the paradox in response to question (ii) above. Consider two populations, [A, B], taken to be mutually exclusive and jointly exhaustive. The measured overall rates for each population are called, [α, β], respectively. Each population is partitioned into categories called, [1, 2], and the measured rates within each partition are called [A_1, A_2, B_1, B_2]. Let's assume that f_1 = the number of females accepted in D_1; F_1 = the total number of females applied to D_1; m_1 = the number of males accepted in D_1; M_1 = the total number of males applied to D_1. Then $A_1 = f_1 / F_1$, and $B_1 = m_1/M_1$. Similarly, we define A_2 and B_2. Let's assume that f_2 = the number of females accepted in D_2; F_2 = the total number of females applied to D_2; m_2 = the number of males accepted in D_2; and M_2 = the total number of males applied to D_2. So, $A_2 = f_2/F_2$ and $B_2 = m_2/M_2$. Likewise, we understand α and β to represent overall rates for each population, females and males, respectively. So the terms $\alpha = \frac{(f_1 + f_2)}{(F_1 + F_2)}$ and $\beta = \frac{(m_1 + m_2)}{(M_1 + M_2)}$. To help conceptualize these notations in terms of Table 1, we provide their corresponding numerical values. $A_1 = \frac{180}{200} = 90\%$, $A_2 = \frac{100}{300} = 33\%$, $B_1 = \frac{480}{600} = 80\%$, $B_2 = \frac{10}{100} = 10\%$, $\alpha, = \frac{280}{500} = 56\%$,

[3]This section we base on our earlier work [reference removed for the sake of review].

and finally $\beta = \frac{490}{700} = 70\%$. Because α, β, A_1, A_2, B_1, and B_2 are rates of some form, they will range between 0 and 1 inclusive. We further stipulate the following definitions.

$C_1 \equiv A_1 \geq B_1$
$C_2 \equiv A_2 \geq B_2$
$C_3 \equiv \beta \geq \alpha$. We call $\mathbf{C} \equiv (C_1 \,\&\, C_2 \,\&\, C_3)$.

For Table 1, C_1 is true because $90\% \geq 80\%$; C_2 is true because $33\% \geq 10\%$, and finally, C_3 is true because $70\% \geq 56\%$. We define the term θ, which provides a connection between the acceptance rates (A_1, B_1, A_2 and B_2) within each partition to their overall acceptance rates (α and β).

$$\theta = (A_1 - B_1) + (A_2 - B_2) + (\beta - \alpha).$$

This condition says for the data in Table 1 that $\theta = 10\% + 23\% + 14\% = 47\%$, meeting the other formal condition for the paradox. That is, Simpson's paradox (SP) arises if and only if

(i) $\mathbf{C} \equiv (C_1 \,\&\, C_2 \,\&\, C_3)$ and
(ii) $\mathbf{C_4} \equiv \theta = \{(A_1 - B_1) + (A_2 - B_2) + (\beta - \alpha)\} > 0$.

Each condition (i and ii) is necessary, but jointly they constitute sufficient conditions for generating SP. This just means that not only Table I satisfies both of these conditions, but also any version of the paradox must satisfy them.

There are three points worth mentioning. First, the characterization of the puzzle in terms of our two conditions captures the central intuitions at stake in the examples given; they are in no way *ad hoc*. The central intuitions are, once again, the reversal or the cessation of an association in the overall population. Second, the paradox is "structural"

in character, in the sense that the reasoning that leads to it is deductive. (Consider our examples, which involve simple arithmetic. The overall rates of acceptance for both females and males follow from their rates of acceptance in two departments taken separately.) Third, unless someone uses the notion of causation trivially, for example, believes that 2+2 "causes" 4, there is no reason to assume that there are causal intuitions lurking in the background. We will return to the last point in greater detail in the following sections.

We now provide an explanation of how the paradox arises in our type I version and why people find it perplexing. For our purposes, we have reconstructed our type I version of SP in terms of its premises and conclusion. However, the point of the reconstruction will be adequately general to be applicable to all types of SP. Before the reconstruction, we introduce a numerical principle called the collapsibility principle (CP) which plays a crucial role in the reconstruction. We call a dataset collapsible if and only if [$A_1 \geq B_1$ and $A_2 \geq B_2$] $\rightarrow \alpha \geq \beta$. Here, the forward arrow "$\rightarrow$" stands for "material implication." The CP says when, if certain relationships hold in the sub-populations between variables (for example, if the rate of acceptance of females is higher than the rate of acceptance of males in both sub-populations), the same relationships must hold in the overall population (that is, the rate of acceptance of females must be higher than the rate of acceptance of males in the population). We will, however, find that the principle in question is, in fact, false with regard to the paradox.

Recall, A_1 and A_2 stand for the rates of acceptance for population A in departments 1 and 2 respectively. Similarly, B_1 and B_2 stand for the rates of acceptance for population B in departments 1 and 2 respectively. In contrast, α and β are rates of acceptance for A

and B populations in the overall school. More explicitly, if we use our earlier notations of f_1, F_2, m_1, M_2, then CP implies $[(f_1/F_1) > (m_1/M_2)$ & $(f_2/F_2) > (m_2/M_2)] \rightarrow \left(f_1 + f_2 \big/ F_1 + F_2\right) > \left(m_1 + m_2 \big/ M_1 + M_2\right)$. In the type I version of SP outlined above, even though the data set satisfies the antecedent, that is, A_1 (i.e., f_1/F_2) > B_1 (i.e., m_1/M_1) and A_2 (i.e., f_2/F_2) > B_2 (i.e., m_2/M_2), its consequent remains unsatisfied. As we can see, CP is a numerical inference principle devoid of any causal intuition.

Here is the reconstruction of the type I version.

P1: Female and male populations are mutually exclusive and jointly exhaustive; one can't be a student of both departments along with satisfying two conditions (i & ii) in our characterization of what is called SP.
P2: The acceptance rate of females is higher than that of males in department # 1.
P3: The acceptance rate of females is higher than that of males in department # 2.
P4: If P2 & P3 are true, then the acceptance rate for females is higher than that of males overall.
P5: However, fewer females are admitted overall. (That is, the consequent of P4 becomes false.)

Conclusion: the deductive consequence of P2, P3, P4 and P5 contradict one another; there is a genuine paradox involved.

In our derivation of the paradox, premise 4 plays a crucial role. In our type I version, the rates of acceptance for females are greater than those of males in each department. That is, $A_1 > B_1$ and $A_2 > B_2$, but $\alpha < \beta$. Thus, CP becomes false. In fact, that CP is not generally true is shown by our derivation of a contradiction.

Our answer to the first question, (i), then, is simply that humans tend to invoke CP uncritically, as a rule of thumb, and thereby make mistakes in certain cases about proportions and ratios; they find it paradoxical when their usual expectation that CP is applicable across the board, turns out to be incorrect.

3. The causal account of SP:

Peter Spirtes, Clark Glymour and Richard Scheines (2000) have developed a subject matter-neutral automated causal inference engine that provides causal relationships among variables from observational data using information about their probabilistic correlations and assumptions about their causal structure. These assumptions are, (i) the Causal Markov Condition (CMC), (ii) the Faithfulness Condition (FC) and (iii) the Causal Sufficiency Condition (CSC). According to CMC, a variable X is independent of every other variable (except X's effects) conditional on all of its direct causes. A is a direct cause of X if A exerts a causal influence on X that is not mediated by any other variables in a given graph. The FC says that all the conditional independencies in the graph are only implied by CMC, while CSC states that all common causes of measured variables are explicitly included in the model. Since these theorists are interested in teasing out reliable causal relationships from data they would like to make sure that those probability distributions are faithful, otherwise, they will not be able to derive causal relationships. In Table 2 above, the dependency we observe between "gender" and "acceptance rate" (in the subpopulation) gets cancelled out by their independence (from the overall population). In this case, the CMC alone imposes no constraints on the distributions that this structure could produce, since there is no independence whatsoever from using CMC. If there is an independence relation in the population that is not a consequence of the CMC, then the population, according to these causal theorists, is unfaithful. By assuming FC, they are able to eliminate all such cases of SP from consideration.

One reason for SP being causal, according to this account, is that (per our first example), applying to the school is a causal problem involving causal dependencies between "gender" and "acceptance rates." Similarly, with regard to Simpson's own example in the literature, Spirtes et al. write, "[t]he question is what *causal dependencies* can produce such a table, and that question is properly known as "Simpson's paradox". (Spirtes, Glymour, and Scheines, 2000, p.40, emphasis is added.)" Therefore, Simpson's worry has a causal story, i.e., the source of the paradox lies in its causal root.

Consider the following two tables to see what these theorists mean. Table III is based on data for 80 patients. 40 patients were given treatment T and 40 assigned to a control, ~T. Patients either recovered, R, or didn't recover, ~R. There were two types of patients, (i) males (M) and (ii) females (~M).

Simpson's Paradox (Medical Example)

Two Groups	M		~M		Recovery Rates		Overall Recovery Rates
	R	~R	R	~R	M	~M	
T	18	12	2	8	60%	20%	50%
~T	7	3	9	21	70%	30%	40%

Table 3

One would think that treatment is preferable to control in the combined statistics, whereas, given the statistics of the sub-population, one gathers the impression that control is better for both men and women. Given a person of unknown sex, would one recommend the control? Spirtes et al. recommend "control." Call this first example the medical example. In a second example, however, we are asked to consider the same data, but now regarding varieties of plants (white [W] or black variety [~W]), R and ~R as yields (high[Y] or low yield [~Y]) and M and ~M as tall and short plants ([T] or [~T]).

Simpson's Paradox (Agricultural Example)

Two Groups	T		~T		Yield Rates		Overall Yield Rates
	Y	~Y	Y	~Y	T	~T	
W	18	12	2	8	60%	20%	50%
~W	7	3	9	21	70%	30%	40%

Table 4

Given Table 4, the overall yield rate suggests that planting the white variety is preferable since it is 10% better overall, although the white variety is 10% worse among both tall and short plants (sub-population statistics). Which statistics should one follow in choosing between which varieties to plant in the future? The CMU theorists' recommendation is that in this case one should take the combined statistics and thus recommend the white variety for planting which is in stark contrast to the recommendation given in the medical case. In short, both medical and agricultural examples provide varying responses to the "what to do question?" There is no *unique* response regarding which statistics, subpopulation or whole, to follow in every case of SP.

Consider the "causal feature" in their causal account concerning the medical example. The novelty of their approach exploits the idea of intervention with regard to these cases. They construe "interventions" as something which directly controls targeted manipulated variables in such a manner that makes the manipulated variables probabilistically independent of all their other causes when the rest of casual structure remains intact. Thus, gender turns out not to be an effect of treatment. When we "intervene" in their technical sense to impose a treatment on a new subject, gender and treatment not only as are but must be probabilistically independent. The reason for

treating gender and treatment to be independent is that we don't know the new subject's gender, leaving no effect on our choice concerning the value of that individual in the casual structure. In the medical example, Spirtes et al. recommend "control."` In the agricultural example, by contrast, the decision to plant which variety does not influence the genetic features in terms of both their association between height and color, and other possible effects on the plant. So whatever causal dependency there is between the plants' height and color in the sample will continue to exist in the population. Thus, recommending the whole population statistics, according to them, should be its natural choice.

4. A counterexample to the causal account:

It is not easy to come up with an example which precludes invoking some sort of appeal to "causal intuitions" with regard to SP. But what follows is, we think, such a case. It tests in a crucial way the persuasiveness of the CMU theorists' account.

Suppose we have two bags of marbles, all of which are either big or small, and red or blue. Suppose in each bag, the proportion of big marbles that are red is greater than the portion of small marbles that are red. Now suppose we pour all the marbles from both bags into a box. Would we expect the portion of big marbles in the box that are red to be greater than the portion of small marbles in the box that are red? Most of us would be surprised to find that our usual expectation is incorrect. The big marbles in bag 1 have a higher ratio of red to blue marbles than do the small marbles; the same is true about the ratio in bag 2. But considering all the marbles together, the small marbles have a higher ratio of reds to blues than the big marbles do. To help understand this scenario, we provide the following example.

Simpson's Paradox (Marble Example)

Marbles of two sizes	Bag 1		Bag. 2		Red marbles rates		Overall rates for red marbles
	Red	Blue	Red	Blue	Bag 1	Bag 2	
Big marbles	180	20	100	200	90%	33%	56%
Small marbles	480	120	10	90	80%	10%	70%

Table 5

In Table 5, we find that in both bags, big marbles have a higher ratio of red to blue marbles than the small marbles do (bag 1: 90% > 80% and bag 2: 33% >10%). When all marbles are pooled together in one bag, the small marbles, however, have a higher ratio of red to blue marbles than do the big marbles (in the combined bag: 70% > 56%). We argue that this is a case of SP since it has the same mathematical structure as the type I version of Simpson's paradox. There are no causal assumptions made in this example, no possible causal "confounding." But it still seems surprising. That is the point of the test case. We believe the test case shows that at least sometimes there is a purely mathematical mistake about ratios that people customarily make. Some statisticians might be tempted to contend that even in this situation there is confounding between the effects of the marble size on the color with the effects of the bag on the color. However, this confounding is not a causal confounding on which the causal account rests since one cannot say that bag 1 has caused big marbles to become more likely to be red or that bag 2 has caused big marbles to become more likely to be blue. In short, one must admit that the above counter-example does not involve causal intuitions, yet it is still a case of SP.

It must also be admitted that there are all sorts of complexities about going from correlation to causation. Correlations are not causes, though correlations are part of the evidence for causes. But what is paradoxical about SP has little to do with these complexities; there is simply a mistaken inference about correlations, which are really

just ratios. Of course, when there are different correlations available which may seem to support conflicting causal inferences, the inference from correlations to cause becomes much more difficult; no one could reasonably deny that. We certainly admit that surprising facts about proportions come up frequently when we infer causes from proportions. This is when our mistakes about proportions seem most troubling to us. But the paradoxical nature of the examples really lies in the mistaken assumptions about the correlations (ratios) themselves.

5. Comparison

We contend that whether SP has anything to do with causality depends on which question (noted above) we are asking. Although first two questions are no doubt distinct, our formal reconstruction of the paradox provides a unified account of them, which empirical studies we have carried out both illustrate and amplify We now discuss the results of two additional experiments. One involves a version of the paradox in non-mathematical language and the second one is in mathematical language. The purpose of this set of experiments is to determine student responses to the questions below. The non-mathematically explained case of the paradox is: [4]

[4]Here, we are overlooking various subtleties involved in setting up those experiments. We offered two separate pages to each student at two different times during a one-hour class period, and two students sitting next to each other were given two different pages which contained the same target questions, but in a different order. We did not want students to know what we were planning to test, nor did we want each student to know exactly what the student sitting next to him/her was doing. Often many of the survey questions are irrelevant to the target questions. For example, we asked, "is there life in Mars"? For fuller versions of those two experiments administered to the students during their class hours please contact the authors.

There are only two high schools in a certain school district. Given that the graduation rate for girls in School #1 is higher than the graduation rate for boys in School 1, and that the graduation rate for girls in School #2 is higher than the graduation rate for boys in School 2. Does it follow that the graduation rate for girls in the district is higher than the graduation rate for boys in the district?

Which one of the following is true?

a. Yes, the graduation rate for girls is **greater than** it is for boys in the district.
b. No, the graduation rate for girls is **less than** it is for boys in the district.
c. No; the graduation rates for girls and boys are **equal** in the district
d. No inference could be made about the truth or falsity of the above because there is not enough information.

The mathematical case of the paradox is:

1. (f1/F1) > (m1/M1).
2. (f2/F2) > (m2/M2).
3. Does it follow that $\left(f_1 + f_2 \middle/ F_1 + F_2 \right) > \left(m_1 + m_2 \middle/ M_1 + M_2 \right)$?

Which one of the following is true?

(a) Yes, the first expression is **greater than** the second.
(b) No, the first expression is **less than** the second.
(c) No, the first and second expressions are **equal**.
(d) No, inference could be made about the truth or falsity of the above because there is not enough information.

The correct answer to both questions is (d). Data were collected from 106 students (n). We found that for the non-mathematical question, students chose response (a) 83% of the time which involves the mistaken use of the collapsibility principle which is a non-causal numerical inference principle. They correctly responded choosing (d) only 12% of the time. For the mathematical question, they are right at the rate of 29%, whereas they have committed the error at 57% of the time. A test of the null hypothesis of no difference in the rate of errors between the two versions of the questions produces evidence of a statistically significant difference in the error rates (P-value < 0.0001). This just means

that the two types of questions produce different error rates.[5] Similar surveys over many years of students in philosophy classes have manifested the same patterns of responses. The varying error-rate in two types of questions is clearly evidence for the statistical difference between them without implying that this statistical difference has any deep philosophical bearing on our discussion, since a large number of students committed the same type of error by misapplying the collapsibility principle in both cases.

The math version of the paradox exactly mirrors our test case which does not involve any causal intuition whatsoever. In turn, the math version also has similar structure as the non-math version of our experiment involving the paradox. Consequently, it will be a mistake to think that the subjects' responses have exploited a causal intuition underlying different versions of the paradox based on the reason that there is no difference between these two experiments, as they exhibit the similar mathematical structure. Most subjects mistakenly applied the non-causal principle CP.

Consider Spirtes, Glymour and Scheines' comments that Simpson's worry has a causal story, i.e., whether to recommend "treatment" or "leave the patients untreated." They show clearly that the source of their recommendation about SP lies in their causal analysis, especially when they recommend "control" in the medical example using intervention and other causal machinery, and recommend planting "white verities" in the agricultural example, finding the same causal dependency between plants' height and color both in sample and population. Two points need to be mentioned clearly here.

[5]Why the students provided a different type of response with regard to the math formulation of the paradox could be an interesting topic to speculate. Students may be frightened by mathematics, and when faced with (d) "non inference is possible", might consider it to be an easy alternative. However, this type of speculation goes beyond the scope of the paper.

First, there is no point in denying that there are causal considerations involved in examples, the medicine and agricultural examples. They have no doubt contributed to our understanding regarding how to address the "what to do" question. Doing is almost always causing something to happen or to be the case. So to know what to do, we generally need to know how to cause something. In the agricultural example, if the decision is about what to plant to get the best yield, then it seems that causal issues settle the case. If we have no way of controlling how many tall plants are produced except by choosing whether to plant the white or black variety, and what we are interested in is what strategy will produce the highest yield, then this settles which yield statistics to pay attention to. However, if the decision question is instead about how to develop varieties producing higher yields, then perhaps one would want to focus on the fact that the most important factor to work on is the size of the plants. So for that question, the subgroup statistics would be relevant.

The second point is about our assumption that the causal decision theory is correct. It is hard to see why one would recommend doing something that is merely correlated with a good result if there is no relevant cause underlying the correlation. Given a set of options constituting a decision situation, decision theory recommends an option which maximizes utility. It makes an appraisal of an option's utility by computing that option's expected utility. This account exploits probabilities and utilities of an option's possible outcome to calculate its expected utility. Here, probabilities in question are dependent on the option. What is distinctive about causal decision theory is that it adopts the dependence of probabilities on the option to be causal rather than merely evidential. Since the what-to-do question is a decision theoretic question and we agree with causal theorists that causal

considerations settle the issues, we will assume the recommendations provided by causal decision theory to be the correct with regard to both medical and agricultural examples.

Given what is discussed so far about the causal theorists' stance toward the what-to-do question one realizes that they have in fact addressed the "what-to do-"question. We, however, argue that they fail to provide an adequate response to the first two questions. We have already provided a counter-example showing that SP has nothing to do with causality in so far as the first two questions are concerned. But SP still seems surprising because the CP violation is what causes this "paradoxical" result. In addition to the fact that the casual theorists have not provided an explanation for its surprising nature, they didn't actually provide conditions for the paradox to arise, our second question about the paradox. We don't deny that causal inference plays a crucial role in addressing the "what-to-do" question. In short, Spirtes et al. address the third question, but not the first two questions, thus failing to distinguish the three types of questions with regard to the paradox. The following table summarizes how the two approaches have addressed three types of questions.

Simpson's Paradox and Three Types of Questions

Approaches	*Why paradoxical?*	*What conditions needed for SP?*	*What to do?*
Causal	No explanation provided	No specific conditions provided	Exploits the idea of intervention
Logic-based	The failure of collapsibility principle	Two conditions provided	Agrees with the causal approach supplemented with causal decision theory

Table 6

6. A possible objection to our account:

One objection that has been raised recently against our account is that the real crux of the paradox lies in knowing *why* it has happened rather than *how* to recognize the paradox when it did.[6] According to this objection, causal theorists are interested in the deeper "why" question. The objector even contends that it is not even hard to provide a causal story behind our counterexample.

If one were asked why the ratios of small red to small blue and large red to large reds of the bag of marbles in the hobby store are what they are, one could provide plausible causal explanations. The manufacturing or packaging process might have favored this ratio, for example. The objector continues that perhaps blue marbles are made of more brittle materials, and so break and are defective more often (or the blue material is more expensive and so the manufacturer wanted a good ratio of blue to red marbles loads with small blues and large reds, or so on). Moreover, we will always look for a causal account rather than rest content with a statistical anomaly.

People may make this causal assumption. One needs to be reminded, however, of our original question: what makes the SP paradoxical? But, (a) our claim about the paradoxical nature of SP is independent of our ability to come up with plausible explanations, (b) we have gathered empirical evidence to support our claim that people extend the collapsibility principle across the board, and (c) we have demonstrated how the collapsibility principle in SP cases leads directly to contradiction.

This new "defense" of the causal theorists changes the question. Instead of asking what makes the SP paradoxical, it asks why we got the paradoxical data. We are not

[6] Our APA commentator has suggested this way-out for causal theorists.

committed to denying that there is typically a causal story about how we came up with the data, although it could perfectly well be a coincidence. And indeed, we have investigated how evidence can be assessed for SP to be able to rule out its occurrence just by chance.[7]

7. Conclusion:

In the first epigram for this paper, as the CMU theorists write, "[t]he skeptic about causality pushes the brake pedal to make his car slow, flips a switch to make a lamp glow, puts his money in the bank to collect interest." (Spirtes, Glymour, and Scheines, 2000, p. 2.) In the same vein, Pearl notes, physicists have "continued to write equations in the office and talk cause-effect in the cafeteria. Physicists talk, write, and think one way and formulate physics in another." (Pearl, 2009, pp. 407-8) In the light of the data presented above, how do we make sense of what they claim? Their underlying theme is that many skeptics' theoretical attitudes toward causality do not agree with their psychological attitudes toward it. Yet in a straightforward sense, our experimental data have shown that psychological attitudes concerning causality do not always enter into SP examples. In fact, the generation of the paradox has nothing to do with causality, since the principle of collapsibility is non-causal. The role of causation in explaining what SP is and why it occurs could be left aside, even if causality remains important to resolving both the most common instances of SP and the what -to -do –questions".

We showed that Simpson's paradox can be generated in a straightforward deductive way. Among its premises is concealed a distinctly human dimension. In recent

[7] (Unpublished)

years, there has been a great deal of discussion of human frailty in connection with an individual's assessment of probabilistic statements (Kahneman et al., 1982; Kahneman, 2011). Our resolution of the paradox has illuminated another aspect of human frailty. We explained its apparent paradoxical nature by invoking the failure of our widespread intuitions about numerical inference. The failure of collapsibility, which is non-causal, in Simpson's paradox-type cases is what makes them puzzling, and the latter is what paints a human face onto the rather abstract structure of "Simpson's paradox."

As George Berkeley once observed, philosophers "have first raised a dust and then complain that [they] cannot see" (Berkeley, 1710). Failing to see the relevance of the three types of questions is what we consider to be the dust that revolves round the paradox. Once the dust has settled, we perceive that the two experiments regarding the paradox have even brought a new flavor to doing "experimental philosophy," since it allows us to decide between two competing accounts of the paradox, (i) causal and (ii) non-causal accounts. However, strictly speaking, which account is the correct one depends entirely on which of the three questions we are asking. This oversight, we contend, is the root cause of the debate over the true nature of Simpson's paradox that the causal theorists have sorely missed.

Select Bibliography

1. Berkeley, G (1710): *A Treatise Concerning the Principles of Human Knowledge*. Turbaynes, C.Am., ed., Indianapolis: Boobs-Merrils, 1710. Edition is used by Turbayne's 1970 edition.
2. Blyth, C (1972): "On Simpson's Paradox and the Sure-Thing Principle." *Journal of the American Statistical Association*: Vol. 67, Number, 338: Theory and Method Section, pp. 364-366.
3. Clark, M (2002): *Paradoxes from A to Z*: London, Routledge.
4. Elles, E., and E. Sober (1983): "Probabilistic Causality and the Question of Transitivity." *Philosophy of Science*, Vol. 50; pp.35-57.
5. Freedman, D, R. Pisani and R. Purve (1999): *Statistics*: (3rd edition) W. W. Norton & Company, New York.
6. Good, I. J., and Y. Mittal (1988): "The Amalgamation and Geometry of Two-By-Two Contingency Tables." *The Annals of Statistics*, Vol. 15, No 2, pp. 694-711.
7. Greenland, S., J. M. Robins, and J. Pearl (1999): "Confounding and Collapsibility in Causal Inference." *Statistical Science*, Vol. 19, pp. 29-46.
8. Harper W, Stalnaker, and Pearce, G (1981) *Ifs: Conditionals, Beliefs, Decision, Chance, and Time*. (eds.) Dordrecht: Reidel.
9. Hoover, K (2001): *Causality in Microeconomics*. Cambridge University Press, England.
10. Joyce, J (1999): *Foundations of Causal Decision Theory*. Cambridge University Press.
11. Kahneman, D, P. Slovic, and A. Tversky (eds.) (1982) *Judgment under Uncertainty: Heuristics and Basics*. Cambridge University Press, England.
12. Kahneman, D. (2011) *Thinking Fast and Slow* Farrar, Straus and Giroux, New York.
13. Kyburg, H (1997): "The Rule of Adjunction and Reasonable Inference." *Journal of Philosophy*, VL. XCIV, No 3, March, pp. 109-125.
14. Lindley, D., and M. Novick (1981): "The Role of Exchangeability in Inference." In *Annals of Statistics*, vol. 9, No.1, pp.45-58.
15. Levi, I. (2000): "Review Essay on the *Foundations of Causal Decision Theory* by James Joyce." *Journal of Philosophy*, 97: pp.387-402.
16. Malinas, G (2001): "Simpson's Paradox: A Logically Benign, Empirically Treacherous Hydra." *The Monist*, vol. 84, no 2, pp. 265-283.
17. Meek, C., and C. Glymour (1994): "Conditioning and Intervening." *British Journal for the Philosophy of Science*, Vol. 45, pp. 1001-21.
18. Mittal, Y (1991): "Homogeneity of Subpopulations and Simpson's Paradox." *Journal of the American Statistical Association*." Vol. 86: pp.167-172.
19. Morgan, S, and C. Winship (2007): *Counterfactuals and Causal Inference*. Cambridge University Press. Cambridge.
20. Novick. M. R. (1983): "The Centrality of Lord's Paradox and Exchangeability for all Statistical Inference." In H. Wainer and S. Messick (eds.), *Principles of Modern Psychological Measurement*. Hillsdale, NJ: Erlbaum.
21. Otte, R (1985): "Probabilistic Causality and Simpson's Paradox." *Philosophy of Science*, Vol. 52; no. 1: pp.110-125.

22. Pavlides, M. and M. Perlman (2009) “How Likely is Simpson’s Paradox?” *The American Statistician*, Vol. 63; no. 3: pp. 226-233.
23. Pearl, J (2000 and 2009): *Causality*: Cambridge: Cambridge University Press.
24. Rothman, K, and S. Greenland (1998): *Modern Epidemiology*. Second Edition. Lippincott Williams: Philadelphia.
25. Shadish, W, and T. Cook, and D. Campbell. (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company: Boston.
26. Simpson, H (1951): “The Interpretation of Interaction in Contingency Tables,” *Journal of the Royal Statistical Society*. Series. B, 13, No 2, pp.238-241.
27. Skyrms, B (1980): *Causal Necessity*. Yale University Press, New York.
28. Sloven, Steven (2005): *Causal Models: How People Thinks about the World and its Alternatives*. Oxford University Press. New York
29. Sober, E and Wilson, D (1998); *Unto Others*: *The Evolution and Psychology of Unselfish Behavior*: Mass: Harvard University Press.
30. Spirtes, P., C. Glymour, and R. Scheines (2000): *Causation, Prediction, and Search*. MIT Press, Cambridge.