Truths about Simpson's Paradox: Saving the Paradox from Falsity

Prasanta S. Bandyopadhyay[1], R. Venkata Raghavan[2], Don Wallace Dcruz[2] and Gordon Brittan Jr.[1]

[1]Department of Philosophy, Montana State University, Bozeman, USA
psb@montana.edu, gbrittan17@gmail.com
[2]Department of Philosophy, University of Hyderabad, Hyderabad, India
{raghavan.rv,don.wallace}@uohyd.ac.in

Abstract. There are three questions associated with Simpson's paradox (SP): (i) Why is SP paradoxical? (ii) What conditions generate SP? and (iii) How to proceed when confronted with SP? An adequate analysis of the paradox starts by distinguishing these three questions. Then, by developing a formal account of SP, and substantiating it with a counter-example to causal accounts, we argue that there are no causal factors at play in answering questions (i) and (ii). Causality enters only in connection with action.

Keywords: Simpson's Paradox, Formal Analysis, Collapsibility Principle, Inference rules, Causal Accounts, Definition of Paradox, First-level-truth, Second-level-truth.

1 Overview

In his recent book, *Saving Truth from Paradox*, Hartry Field discusses the philosophical significance of paradoxes. According to him, "[a]ny resolution of the paradoxes will involve giving up (or at least restricting) some very firmly held principles:... [and] [t]he principles to be given up, are the ones to which the average person simply can't conceive of alternatives. That's why the paradoxes are *paradoxes*." [4, p.17]. Their significance and the firmly held principles which we have to give up in resolving them is a recurring theme in philosophical logic. We will illustrate this in the case of Simpson's paradox (SP), which involves the reversal of the direction of a comparison or the cessation of an association when data from several groups are combined to form a single whole [17]. At least *three* distinct questions are important in understanding the nature of the paradox: (i) Why or in what sense, is SP a paradox? (ii) What are the conditions in which the paradox arises? (iii) How should one proceed when confronted with a typical case of the paradox, hereafter to be called the "what-to-do" question?[1] The three questions are distinct: answering one of them does not entail answers to the

[1] Daniel Hausman was perhaps the first philosopher who drew our attention to the significance of these three types of questions (in an email communication).

M. Banerjee and S.N. Krishna (eds.): ICLA 2015, LNCS 8923, pp. 58–73, 2015.

others. Following these three questions, we distinguish two types of truth about SP: the first-level truth and the second-level truth. The significance of the three questions about the paradox is what we call the "first-level truth", while the significance of the first two questions in unlocking its paradoxical nature and the conditions for its emergence is what we call the "second-level truth." The failure to appreciate the difference between these two levels of truth, we will contend, is the source of its misdiagnosis. Tables 1 and 2 illustrate the two types of SP. The data in both tables represent acceptance and rejection rates of male and female applicants for graduate school in two departments of an imaginary university in some year.

Table 1. Simpson's Paradox (Type I)

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

Table 2. 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%	13%	10%

Table 1 represents an example of the paradox in which the association in the sub-populations (Dept 1 and Dept 2) with higher acceptance rate for females is *reversed* in the combined population, with overall higher rates for males. Table 2 is an example that shows the paradoxical effect when the association between "gender" and "acceptance rates" in the sub-populations *ceases* to exist in the combined population. Though the acceptance rates for females are higher in each department, in the combined population, those rates cease to be different.

This paper is divided into eight sections. In section two, we will propose our response to the first two questions. Then we will briefly introduce two influential causal accounts of SP proposed independently by Judea Pearl [9] and Peter Spirtes, Clark Glymour and Richard Scheines (hereafter called 'SGS') [15]. In section four, we will produce a counter-example to the causal accounts. The next section will be devoted to the "what-to-do" question. In section six, we evaluate causal accounts (with special attention to Pearl's) of the paradox and compare them with ours. In section seven, we will discuss how our account affects the general notion of paradoxes and their classification while providing a general definition of a paradox. We conclude with some remarks in section eight.

2 Formal Analysis of SP

2.1 Conditions of SP[2]

We begin with an analysis of the paradox in response to question (ii), "what are the conditions in which the paradox arises?" Consider two groups, [A, B], taken to be mutually exclusive and jointly exhaustive. The overall rates for each group are [α, β] respectively. Each group is partitioned into categories [1, 2] and the rates within each partition are [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 $B_1 = m_1/M_1$. Defining f_2, F_2, m_2 and M_2 in a similar way, we get $A_2 = f_2/F_2$ and $B_2 = m_2/M_2$. Likewise, we could understand α and β as representing the overall rates for females and males, respectively. So the terms $\alpha = (f_1+f_2)/(F_1+F_2)$ and $\beta = (m_1+m_2)/(M_1+M_2)$. To help conceptualize these notations in terms of Table 1, we provide their corresponding numerical values: $A_1 = 180/200 = 90\%$, $A_2 = 100/300 = 33\%$, $B_1 = 480/600 = 80\%$, $B_2 = 10/100 = 10\%$, $\alpha = 280/500 = 56\%$, and finally $\beta = 490/700 = 70\%$. Since 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 where, "$\equiv$" means "is defined as".

$$C_1 \equiv A_1 \geq B_1.$$

$$C_2 \equiv A_2 \geq B_2.$$

$$C_3 \equiv \beta \geq \alpha.$$

$$C \equiv (C_1 \& C_2 \& C_3).$$

In terms of Table 1, these definitions become C_1: $90\% > 80\%$, C_2: $33\% > 10\%$, C_3: $70\% > 56\%$ and thus C is satisfied. But C alone is not a sufficient condition for SP. We could have a case where $A_1 = B_1$, $A_2 = B_2$ and $\beta = \alpha$ resulting in no paradox, yet C being satisfied. Hence, we stipulate another definition:

$$C_4 \equiv \theta > 0.$$

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

For the data in Table 1, θ equals 10% + 23% + 14%. Again, C_4 alone is not sufficient for SP since we could have a case where $A_1 > B_1$, $B_2 > A_2$ and $\beta > \alpha$ resulting in no paradox (C is violated) and yet C_4 being satisfied.[3] Hence,

[2] Some parts of this section are based on our previous work [1,2].

[3] As a heuristic rule we take A_1 to be that sub-group ratio which is the greater of the two ratios and B_1 as that which is the lesser of the two. In table 1, the ratio of women admitted to department 1 is greater than that of men. Hence, the former will be taken as A_1 and the latter will be taken as B_1. Similarly, since the ratio of women admitted to department 2 is greater than that of men, the former is taken as A_2 and the latter as B_2. This avoids the complexity of taking the absolute value of their difference in the calculation of θ.

a situation is a case of SP if and only if:

$$C \equiv (C_1 \& C_2 \& C_3) \tag{a}$$

$$C_4 \equiv \theta = (A_1 - B_1) + (A_2 - B_2) + (\beta - \alpha) > 0^4 \tag{b}$$

Both (a) and (b) are necessary conditions, but they jointly constitute sufficient conditions for generating SP [1].[5] Both conditions for the paradox generate two key theorems which specify the relationship between the two acceptance rates in both sub-populations. These are: 1. $A_1 \neq A_2$, and 2. $B_1 \neq B_2$. Table 3 shows why the condition for Theorem 1 needs to hold. Since $A_1 = A_2$, i.e., $25\% = 25\%$, no paradox results. Similarly, in Table 4, since $B_1 = B_2$, i.e., $25\% = 25\%$, the paradox does not occur. Proofs of these theorems are provided in the appendix.

Table 3. No SP ($A_1 = A_2$)

Two Groups	Dept 1		Dept 2		Acceptance Rates		Overall Acceptance Rates
	Accept	Reject	Accept	Reject	Dept 1	Dept 2	
Females	75	225	75	225	25%	25%	25%
Males	10	90	20	80	10%	20%	15%

Table 4. No SP ($B_1 = B_2$)

Two Groups	Dept 1		Dept 2		Acceptance Rates		Overall Acceptance Rates
	Accept	Reject	Accept	Reject	Dept 1	Dept 2	
Females	10	90	20	80	10%	20%	15%
Males	75	225	75	225	25%	25%	25%

There are four points worth mentioning. First, Clark Glymour [5] would call our account an application of the "Socratic method" in which we provide necessary and sufficient conditions for the analysis of a concept.[6] Second, the characterization of the puzzle in terms of our two conditions captures the paradoxical

[4] See Blyth [3] for similar conditions. However, our conditions and notations are slightly different from his.

[5] See [6], [16]. The latter paper shows that SP reversal involves Boolean disjunction of events in an algebra rather than being restricted to cells of a partition.

[6] Glymour contrasts this method with what he calls the "Euclidean"-method based theories where one could derive interesting consequences from them although Euclidean-method based theories, according to him, are invariably incomplete. It is interesting to note two very different points. First, although Glymour is not fond of the Socratic-method on which, however, a large part of the western philosophical tradition rests, our Socratic-method based logical account at the same time is also able to generate some interesting logical consequences (See [1,2]). Second, it is not only the Greeks who applied this method. In classical Indian philosophical tradition, the Socratic method is also very much prevalent where a definition of a term is evaluated in terms of whether it is able to escape from being both "too narrow" and "too wide."

nature of the data in the examples given, namely, the reversal or the cessation of an association in the overall population; they are in no way ad hoc. Third, 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. Note that both conditions of the paradox can be defined in terms of the probability theory, which is purely deductive [3]. Fourth, 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.

2.2 Why is SP "Paradoxical"?

To answer question (i), "why is SP a paradox?" we now provide an explanation of how the paradox arises in people's minds and why it is found perplexing. In other words, what is the reasoning that the "average person" follows that leads him/her to a paradoxical conclusion? For our purposes, we have reconstructed our type I version of SP in terms of its premises and conclusion to show how the paradox arises. However, the point of the reconstruction will be adequately general to be applicable to all types of SP. We introduce a numerical principle called the collapsibility principle (CP) which plays a crucial role in the reconstruction. CP says that relationships between variables that hold in the sub-populations (e.g., the rate of acceptance of females being higher than the rate of acceptance of males in both sub-populations) must hold in the overall population as well (i.e., the rate of acceptance of females must be higher than the rate of acceptance of males in the population). There are two versions of CP corresponding to the two types of SP represented by Tables 1 and 2. The first version of CP (CP1) says that a dataset is collapsible if and only if $[(A_1 > B_1)\&(A_2 > B_2) \rightarrow (\alpha > \beta)]$. The second version of CP (CP2) states that a dataset is collapsible if and only if $[(A_1 = B_1)\&(A_1 = B_2) \rightarrow (\alpha = \beta)]$. That CP1 and CP2 can lead to paradoxical results demonstrates that both versions of the principle are not, in all their applications, true. That is, CP $\rightarrow \sim$SP, whether it is CP1 or CP2, where "$\rightarrow$" is to be construed as the implication sign. If f_1, F_2, m_1, M_2, A_1, A_2, B_1, B_2, α, and β have the same meanings as given in section 2.1, then CP1 takes the following form.

$$\left(\left(\frac{f_1}{F_1} > \frac{m_1}{M_1}\right) \& \left(\frac{f_2}{F_2} > \frac{m_2}{M_2}\right)\right) \rightarrow \left(\frac{f_1 + f_2}{F_1 + F_2} > \frac{m_1 + m_2}{M_1 + M_2}\right)$$

Likewise, CP2 says

$$\left(\left(\frac{f_1}{F_1} = \frac{m_1}{M_1}\right) \& \left(\frac{f_2}{F_2} = \frac{m_2}{M_2}\right)\right) \rightarrow \left(\frac{f_1 + f_2}{F_1 + F_2} = \frac{m_1 + m_2}{M_1 + M_2}\right)$$

As we can see, CP is a numerical inference principle devoid of any causal intuition. Here is the reconstruction of type I version of SP:
(1) Female and male populations are mutually exclusive and jointly exhaustive; one can't be a student of both departments and satisfy the two conditions of SP.
(2) The acceptance rate of females is higher than that of males in Department 1. (observed from data)
(3) The acceptance rate of females is higher than that of males in Department 2. (observed from data)
(4) If (2) and (3) are true, then the acceptance rate for females is higher than that of males overall. (from CP1)
(5) Hence the acceptance rate for females is higher than that of males overall. (from (2), (3) and (4))
(6) However, fewer females are admitted overall. (observed from data)
(7) Overall acceptance rate for females is both higher and lower than that of males. (from (5) and (6))
In our derivation of the paradox, premise (4) plays a crucial role. In type I version of SP, as given in Table 1, CP1 does not hold ($A_1 > B_1$ and $A_2 > B_2$, but $\alpha < \beta$). That CP1 is not generally true is shown by our derivation of a contradiction. The same result can be obtained for Type II version of SP in Table 2 where CP2 has to be given up if the paradox is to be avoided.

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. And the reason we think people invoke CP uncritically, is its remarkable (formal) resemblance with the two inference rules given below.[7]
1. In elementary algebra, the following truth holds for real numbers:

$$x_1 > y_1$$
$$x_2 > y_2$$
$$\therefore (x_1 + x_2) > (y_1 + y_2)$$

While it is correct to substitute $A_1(f_1/F_1)$ for x_1, $B_1(m_1/M_1)$ for y_1, $A_2(f_2/F_2)$ for x_2 and $B_2(m_2/M_2)$ for y_2, people might confuse $(x_1 + x_2)$ and $(y_1 + y_2)$ for $\alpha((f_1 + f_2)/(F_1 + F_2))$ and $\beta((m_1 + m_2)/(M_1 + M_2))$ respectively, leading them to think that CP is also a mathematical truth. Thus, mistakes about proportions and ratios could lead the average person to see a superficial resemblance between CP and the above mathematical truth.
2. In propositional logic, the following rule is valid:

$$P1 \rightarrow Q \quad \text{(A)}$$
$$P2 \rightarrow Q \quad \text{(B)}$$
$$\therefore (P1 \vee P2) \rightarrow Q \quad \text{(C)}$$

[7] We are thankful to Joseph Hanna and John G. Bennett for helpful emails on this point.

In our case, let P1 = "A student applies to Department 1", P2 = "A student applies to Department 2" and Q = "The student has greater chance of being accepted, if the gender of the student is female". Now, (A) partially captures the condition $A_1 > B_1$ whereas (B) partially captures $A_2 > B_2$. (C), which reads, "If a student applied to Department 1 or Department 2 then, the student has greater chance of being accepted if the gender of the student is female" resembles the condition $\alpha > \beta$. We do not suggest that propositional logic can capture the essence of the paradox. The reasoning leading to SP involves probabilistic considerations which, unlike propositional logic, is not truth-functional. For example, the probability of a disjunction is not a function of the probability of its disjuncts. Likewise, SP is a weighted average of probabilities, or, in other words, averages of averages. No such concept of weighted average exists in truth-functional logic. The above comparison of CP with a valid propositional rule no more than suggests why people tend to use CP even in cases where it leads to contradiction.

3 Causal Accounts of SP

3.1 Pearl's Account

Pearl argues that the arithmetical inferences in SP seem counter-intuitive only because we commonly make two incompatible assumptions, that causal relationships are governed by the laws of probability and that causal relationships are more *stable* than probabilistic relationships [9, pp. 180, 25]. Once we reject either of these assumptions, and he opts for rejecting the first, the "paradox" is no longer paradoxical. On the other hand, when we fail to distinguish causal from statistical hypotheses, the paradox results.

Pearl makes two basic points. One, SP is to be understood in causal terms for its correct diagnosis. In the type I version, for example, the effect on "acceptance" (A) of the explanatory variable, "gender" (G), is hopelessly mixed up (or "confounded") with the effects on A of the other variable, "department" (D). We are interested in the direct effect of G on A and not an indirect effect by way of another variable like D. His other point is that causal hypotheses, which support counterfactuals, often cannot be analyzed in statistical terms. Suppose we would like to know Bill Clinton's place in US history had he not met Monica Lewinsky. The counterfactual for the causal hypothesis is "Clinton's status in the US history would be different had he not met Monica Lewinsky" [9, p. 34]. However, there is no statistical model one could construct that would provide the joint occurrence of 'Clinton' and 'no Lewinsky'. There simply are no appropriate data, as there are, for instance, in the fair coin-flipping experiments where the model about flipping a coin and data about it are well known.

3.2 SGS Account

Spirtes, Glymour, and Scheines suggest 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: 1. Causal Markov Condition (CMC), 2. Faithfulness Condition (FC) and 3. 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 in representing causal relations in them.

One reason for SP being causal, according to this account, is that (for the example given in Table 1) applying to the school has a causal dimension involving causal dependencies between "gender" and "acceptance rate". More female students chose to apply to the departments where rates of acceptance are significantly lower, *causing* their overall rates of acceptance to be lower in the combined population. 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"." [15, p. 40].

4 Counter-Example 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 causal accounts.[8]

Table 5. Simpson's Paradox (Marble Example)

Marbles of two sizes	Bag 1		Bag 2		Rates of red Marbles		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%

Suppose, as in Table 5, 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 (Bag 1: 90% > 80% and Bag 2: 33% > 10%). 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.

[8] This counter-example is due to John G. Bennett.

The big marbles in the first bag have a higher ratio of red to blue marbles than do the small marbles; the same is true about the ratio in the second bag. But considering all the marbles together, the small marbles have a higher ratio of reds to blues than the big marbles do (in the combined bag: $70\% > 56\%$).

We argue that this marble example is a case of SP since it has the same mathematical structure as the type I version of SP. There are no causal assumptions made in this example, no possible causal "confounding" and yet it seems paradoxical. We believe this counter-example shows that at least sometimes, there is a purely mathematical mistake about ratios that people customarily make. Some causal theorists might be tempted to contend that even in this example 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 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.

5 "What-To-Do" Question and Causal Accounts

In the case of SP, "what-to-do" questions arise when investigators are confronted with choosing between two conflicting statistics. For example, in Table 1, the conflict is between the uncombined statistics of the two departments and their combined statistics. Which one should they use to act? It is evident that many interesting cases of choosing actions arise when we infer causes/patterns from proportions. The standard examples[9] deal with cases in which "what-to-do" questions become preeminent. But it should be clear in what follows that there is no unique response to this sort of question for all cases of the paradox. Consider Table 6 based on data about 80 patients. 40 patients were given the treatment, T, and 40 assigned to a control, ∼T. Patients either recovered, R, or didn't recover, ∼R. There were two types of patients, males (M) and females (∼M).

Table 6. 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%

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 gender, would one recommend the control? The standard response is clear: control is better for a person of unknown gender (since $\Pr(R|\sim T) >$

[9] These recommendations are standard because they are agreed upon by philosophers [8], statisticians, and computer scientists [9].

Pr(R|T)). Call this first example 'the medical example'. In the second example, call it 'the agricultural example', we are asked to consider the same data, but now T and ∼T are replaced by the varieties of plants (white [W] or black variety [∼W]), R and ∼R by the yield (high [Y] or low yield [∼Y]) and M and ∼M by the height of plants (tall [T] or short [∼T]).

Table 7. 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%

Given this new interpretation, 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 standard recommendation is to take the combined statistics and thus recommend the white variety for planting (since Pr(Y|W) > Pr(Y|∼W)), which is in stark contrast with the recommendation given in the medical example. 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. We agree with standard recommendations with a proviso, i.e., we need to use substantial background information, which is largely causal in nature, to answer "what-to-do" questions, as *doing* something means *causing* something to happen.

6 Truths about SP: An Evaluation of Causal Accounts

We argued that to understand the significance of SP as a whole, we need to distinguish three types of questions (first-level truth) as well as divorce the first two questions from the third to show that causality is irrelevant both in unlocking the paradoxical nature of SP and providing conditions for its emergence (second-level truth). Based on our discussion of the causal accounts, one realizes that causal theorists have in fact addressed the "what-to-do" question. We don't deny that causal inference plays a crucial role in choosing the right statistic when confronted with the paradox. Hence we agree with both Pearl and SGS about the third question. However, as far as we know, SGS have not distinguished the three questions about SP, and thereby failed to appreciate the first-level truth about SP. Pearl on the other hand, does distinguish the three questions. But both causal accounts fail to understand the second-level truth about the paradox. Notice that one may, like Pearl, recognize the first-level truth and yet fail to recognize the second-level truth. An examination of his responses to the first two questions will reveal the reason behind this, showing how his causal account

falls short of providing an adequate explanation for the first two questions and thereby not being able to appreciate the full significance of SP.

In response to the first question, Pearl draws attention to the distinction between what he calls "Simpson's reversal", which is merely an "arithmetic phenomenon in the calculus of proportions" and "Simpson's paradox" which is "a psychological phenomenon that evokes surprise and disbelief" [10, p. 9]. He thinks that the latter is the result of intuitions guided by causal considerations and the fallacy of equating correlation with causation. While agreeing with him about the fallacy, we pointed out, with the help of the marble counter-example, that fundamentally, SP is devoid of any causal intuitions, although most day-to-day examples of SP can be interpreted causally. We think that human puzzlement about SP stems from the unexpected failure of CP which closely resembles valid inference rules (section 2.2). With respect to the second question, Pearl identifies "scenarios" in which one can expect a reversal. A scenario, according to him, is "a process by which data is generated" [10, p. 10]. The causal calculus/models which represent these causal scenarios are different from our formal conditions which have been derived from the structure of the paradox (section 2.1). So our conditions capture *all* cases of SP regardless of the causal process involved and provide a more general account than either of the causal accounts.

7 Re-evaluating the Place of SP in Paradox Literature

Logicians tend to hold different views concerning what paradoxes are. Whether SP is a paradox depends on how one defines and slices paradoxes. Priest [11], for example, may not consider SP to be a paradox as it is neither a set-theoretic paradox such as Russell's nor a semantic one like the Liar Paradox. But, under Sainsbury's construal, SP could be regarded as a paradox since he understands a paradox as "an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises." [14, p. 1]. However, this might not furnish a genuine rationale for what makes paradoxes paradoxical since one might worry what an "apparently acceptable reasoning" is. In this regard, we find a better explanation in W.V.Quine, who both defines and provides a general rationale for the apparently paradoxical nature of paradoxes. A paradox, according to him, is "just any conclusion that at first sounds absurd but that has an argument to sustain it" [12, p. 1]. SP can be treated as a paradox in this Quinean sense.

Two points are to be noted here. First, Quine's use of the word "absurd" could be ambiguous since it lends itself to two interpretations: a) psychological confusion and b) logical contradiction. Our analysis of SP suggests that SP "sounds absurd" under both interpretations. Given the logical reconstruction of SP (section 2.2), we see how it leads to a self-contradictory conclusion. And, given our response to question (i), we find that people tend to apply CP across the board and their psychological confusion results when they find out that CP, in fact, cannot be so applied. Second, our research shows that the sharp distinction Quine draws between "veridical paradox" and "falsidical paradox"

does not necessarily hold about SP. Distinguishing between these two varieties while justifying each, he writes, "[a] veridical paradox packs a surprise, but the surprise quickly dissipates itself as we ponder the *proof*. A falsidical paradox packs a surprise, but it is seen as a false alarm when we solve the *underlying fallacy*." [12, p. 9, emphasis is ours]. He argues that Gödel's discovery and other paradoxes in set theory are veridical paradoxes. We think that SP can be seen as a case of veridical paradox as soon as we realize that the population data in all tables follow necessarily from the sub-population data along with the proofs we provided for the paradox to hold. To explore whether SP could fall under the category of a falsidical paradox, consider Quine's own example of the latter. According to him, paradoxes of Zeno are instances of falsidical paradoxes since they rest on the fallacious assumption that "an infinite succession of intervals must add up to an infinite interval." Once we note this, it becomes clear that the initial surprise about them was unwarranted. The same reasoning can be offered for SP being a falsidical paradox. Our analysis shows that the surprise SP packs rests on holding the dubious assumption that CP is unconditionally applicable. Once we realize this, the paradoxical nature of SP disappears. So, the unique feature of SP is that it is a paradox in both veridical and falsidical senses. Therefore, there need not be a sharp distinction between these two types of paradoxes as Quine once argued.

Two issues emerge from the preceding discussion. First, we rely on Quine's definition of a paradox and how it fares with regard to SP; As we will see in a moment Roy Sorensen thinks that Quine's definition is flawed as, according to him, it is neither necessary nor sufficient [13]. Second, whether it is possible to advance a definition of a paradox which could include all types of paradoxes including SP and the Liar paradox under its banner. The rest of this section will be devoted to addressing these two issues.

Sorensen's method is to turn the definition of a paradox against what he takes to be Quine's own "paradox" of radical translation. Quine sets out his "paradox" by first assuming the possibility of a "radical translation" situation, in which neither speaker knows a word of the other's language. Consider a group of linguists interested in understanding what the native speakers' utterances mean. Suppose the speakers utter "gavagai." The linguists observe the speakers, hear what they utter, observe the conditions under which they utter a word or sentence, and determine what they are looking at or pointing out when they utter. Armed with such information, let's assume these linguists make a hypothesis that "gavagai" means "rabbit". In the same way, it is possible that another group of linguists having the same evidence as the first group translates "gavagai" as "undetached rabbit part." Which one is the correct translation of "gavagai"? Based on this thought experiment, Quine contends that radical translation is not possible as meaning, here understood as referent, is indeterminate or at least undermined by the totality of empirical evidence that is available. There is no way to know whether the translation of "gavagai" as "rabbit" or "undetached rabbit part" is the correct hypothesis. But the conclusion seems absurd; at least most of the time, we know what others in our language group (or outside it) are referring

to when they utter sounds. Sorensen rejects Quine's construal of paradoxes by pointing out that, " 'What is the translation of 'Gavagai'?' has infinitely many rival answers. According to Quine, the problem is that infinitely many of these are equally good answers. Quine's paradox of radical translation is a counterexample to his own definition of paradox. In addition to showing that absurdity is inessential to paradox, the paradox of radical translation shows that the paradox can be free of arguments and conclusions. 'What is the translation of 'Gavagai'?' has answers obtained by translation, not conclusions derived by arguments."[13, p. 560].

Even though we agree with Sorensen that actual, in contrast to merely apparent, absurdity is not necessary for understanding the nature of paradoxes, we disagree with his claim that it is not helpful to construe paradoxes in terms of an argument consisting of premises and a conclusion. What Sorensen misses in his criticism of Quine is that while a paradox need not present itself in canonical forms, their canonical forms are useful tools in understanding them, just as the canonical form of an argument (with numbered premises and designated conclusion) is a useful tool for discussing arguments that, in real life, do not always present themselves in that way. To force the paradox into the canonical form, suppressed premises must be revealed and hidden assumptions made explicit. If the radical translation claims are paradoxical, they can be fitted into the canonical forms, though there may be different ways to do that. Here's one version in our favored canonical form:

(1) A correct translation of one natural language into another is one that is entirely compatible with all the facts about usage.
(2) If two translations translate a given term in one language into incompatible terms in another language, one of the translations is not correct.
(3) There are two correct translations of the native language word "gavagai" into English; one translates it as "rabbit" and the other translates it as "undetached rabbit part."
(4) "Rabbit" and "undetached rabbit part" are incompatible terms in English (in the sense that they do not have the same referent).
(5) The native language and English are natural languages.

Contrary to Sorensen, we find that it is possible to exhibit the paradox of radical translation in terms of an argument with premises and a conclusion, revealing the assumption on which it rests. At the same time, we agree with Sorensen in a different way when he holds that a paradox need not have a genuinely absurd conclusion. We tend to think that "sounding absurd" lends a psychological air to the issue of a paradox. In light of these two considerations, we propose a general definition of a paradox. A paradox is an (apparently) inconsistent set of sentences each of which seems to be true.[10] The word "apparently" in this account, as in Quine's, is to allow for cases that depend on fallacious arguments, as in the well-known "proofs" for 1=2. Another advantage of this account is that one might make several arguments from a set of inconsistent sentences, but one

[10] We owe this definition to John G. Bennett. Lycan [7] has also provided a similar definition.

would probably not want to call them distinct paradoxes. Any paradox worth the name, including SP, should obey this definition. Simplifying our reconstruction of SP as a paradox in section 2.2, we provide a rough schema for SP with the (optional) false premise marked by an asterisk, for the type 1 version of the paradox.

(1) Sub-population 1 has a positive correlation between two variables.
(2) Sub-population 2 has a positive correlation between two variables.
(*3) If each sub-population in a partition of a larger population exhibits a positive correlation between two variables, then the population as a whole will also exhibit that same positive correlation between the same two variables
(4) Overall population has a negative correlation between the same two variables.

If *3 is included, the set is inconsistent, since premise *3 is false. If *3 is not included, the set seems to be inconsistent, but is not. Whether to analyze the paradox one way or the other may depend on the example and the context. We think that our definition is adequately general to include even the Liar paradox. Call "this sentence is false" the liar sentence. The following provides a canonical reconstruction of the Liar paradox with two premises and a conclusion.

(1) The liar sentence is true.
(2) The liar sentence is false.
(3) A sentence is either true or false, but not both.

In this section, among other issues, we both discussed and evaluated different views on paradoxes. As a result, we are able to provide a general framework to understand paradoxes while showing that both SP and the Liar paradox satisfy it even though the former has an apparently contradictory conclusion while the latter has a genuinely contradictory one.

8 Conclusion

Unraveling paradoxes is crucial to philosophers of logic as they challenge our deeply held intuitions in a fundamental way. While addressing SP, we distinguished three types of questions. We showed that answering one does not necessarily lead to the answers of the rest. Although, admittedly, the "what-to-do" question is the most important insofar as the practical side of SP is concerned, some causal theorists have overlooked the need to distinguish these three questions, thus failing to appreciate the first-level truth about the paradox. Even if they recognize this first-level truth, the importance of the "what-to-do" question drives them to assume that the causal calculus needed to address this question is the correct way to unlock the riddle about the paradox. We, however, showed that the truth about the paradoxical nature of SP and conditions for its emergence need to be isolated from the "what-to-do" question. This failure on the part of the causal theorists leads to their failure in appreciating the second-level truth about the paradox. Pivoting on the question "why is SP paradoxical?", we provide a general framework for understanding any paradox. Our analysis of

SP also highlights the significant role played by CP in generating the paradoxical result. Such principles are what Field would suggest we jettison to escape paradoxes.

Acknowledgments. We wish to thank Prajit Basu, Abhijit Dasgupta, S. G. Kulkarni, Vineet Nair, Davin Nelson and members of the Research Scholar Group of the University of Hyderabad, Philosophy Department (where an earlier version of the paper was presented) and three anonymous referees for their very helpful comments. We are indebted to John G. Bennett for several key email communications concerning the paradox.

References

1. Bandyopadhyay, P.S., Nelson, D., Greenwood, M., Brittan, G., Berwald, J.: The logic of Simpson's paradox. Synthese 181, 185–208 (2011)
2. Bandyopadhyay, P.S., Greenwood, M., Dcruz, Don Wallace F., Venkata Raghavan, R.: Simpson's Paradox and Causality. Am Philos Quart. (forthcoming)
3. Blyth, C.: On Simpson's Paradox and the Sure-Thing Principle. J. Am. Stat. Assoc. 67(338), 364–366 (1972)
4. Field, H.: Saving Truth from Paradox. Oxford University Press, Oxford (2009)
5. Glymour, C.: Critical Notice. James Woodward's Making Things Happen: A Theory of Causal Explanation. Brit. J. Philos. Sci. 55, 779–790 (2004)
6. Good, I., Mitttal, Y.: The amalgamation and geometry of two-by-two contingency tables. Annals of Statistics 15(2), 694–711 (1988)
7. Lycan, W.G.: What, exactly, is a paradox? Analysis 70(4), 615–622 (2010)
8. Meek, C., Glymour, C.: Conditioning and Intervening. Brit. J. Philos. Sci. 45, 1001–1021 (1994)
9. Pearl, J.: Causality. Cambridge University Press, Cambridge (2009)
10. Pearl, J.: Comment: Understanding Simpson's Paradox. Am. Stat. 68(1), 8–13 (2014)
11. Priest, G.: The logic of paradox. J. Philos. Logic. 8(1), 219–241 (1979)
12. Quine, W.V.O.: The Ways of Paradox and Other Essays. Revised and Enlarged. Harvard University Press, Cambridge (1976)
13. Sorensen, R.: A Brief History of the Paradox. Oxford University Press (2003)
14. Sainsbury, R.M.: Paradoxes, 3rd edn. Cambridge University Press, Cambridge (2009)
15. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. MIT Press, Cambridge (2000)
16. Wheeler, G.: Two Puzzles concerning measures of uncertainty and the positive Boolean connectives. In: Neves, J., Santos, M.F., Machado, J.M. (eds.) EPIA 2007. LNCS (LNAI), vol. 4874, pp. 170–180. Springer, Heidelberg (2007)
17. Yule, G.U.: Notes on the Theory of Association of Attributes in Statistics. Biometrika 2(2), 121–134 (1903)

Appendix[11]

For proving theorems 1 and 2 we firstly assume that the conditions of SP (arrived at in section 2.1) are satisfied. That is,

$$C \equiv (C_1 \& C_2 \& C_3)$$
$$\theta = (A_1 - B_1) + (A_2 - B_2) + (\beta - \alpha) > 0$$

Further, we stipulate the following definitions:

$$\text{a} = (\text{members of A in partition 1})/(\text{total members of A})$$
$$\text{b} = (\text{members of B in partition 1})/(\text{total members of B})$$
$$\alpha = aA_1 + A_2(1-a)$$
$$\beta = bB_1 + B_2(1-b)$$

A_1, A_2, B_1 and B_2 have the same meanings defined in section 2.1. We have defined α and β differently than what we had done in section 2.1 only to ease the proofs of the following theorems; otherwise the two sets of definitions are mathematically equivalent. To take an example, in Table 1 (Type I SP), $A_1 = 180/200$, $A_2 = 100/300$, $B_1 = 480/600$, $B_2 = 10/100$, $a = 200/500$, $b = 600/700$. Hence, $\alpha = (180/500) + (100/500) = 280/500 = 56\%$ and $\beta = (480/700) + (10/700) = 490/700 = 70\%$.

Theorem 1. *Simpson's paradox results only if* $A_1 \neq A_2$.

Proof: Let us assume that $A_1 = A_2$. Then, $\alpha = aA_1 + A_2(1-a) = aA_1 + A_2 - aA_2 = A_1 = A_2$. Given this, there are three possible scenarios. (I) $B_1 > B_2$, or (II) $B_1 < B_2$ or (III) $B_1 = B_2$.

(I) If $B_1 > B_2$, then $[B_1 b + B_1(1-b)] > [B_1 b + B_2(1-b)]$. Therefore, $B_1 > \beta$. Yet, if $A_1 \geq B_1$, and $\alpha = A_1$, then $\alpha > \beta$, which contradicts the assumption that $\beta \geq \alpha$. Therefore, if $A_1 = A_2$, then it can't be that $B_1 > B_2$.

(II) If $B_1 < B_2$, then $[B_1 b + B_2(1-b) < [B_2 b + B_2(1-b)] = B_2$. Therefore, $\beta < B_2$. Yet, $A_2 \geq B_2$, $A_1 \geq B_2$, and $\alpha \geq B_2 > \beta$. This contradicts the assumption that $\beta \geq \alpha$. Therefore, if $A_1 = A_2$, then it can't be the case that $B_1 < B_2$.

(III) If $B_1 = B_2$, then $\beta = bB_1 + B_2(1-b) = bB_1 + B_1(1-b) = B_1$. Given that $A_1 \geq B_1$, $A_1 = \alpha$, and $B_1 = \beta$, then $\alpha \geq \beta$. Yet, by assumption, $\beta \geq \alpha$. Therefore, $\beta = \alpha$. Since $A_1 = A_2 = \alpha$, and $B_1 = B_2 = \beta$, it must be that $A_1 = B_1$, $A_2 = B_2$, and $\alpha = \beta$. That $\alpha = \beta$ contradicts the assumption that out case is paradoxical, characterized by the reversal which we don't find here. Therefore, if $A_1 = A_2$, it can't be the case that $B_1 = B_2$. Therefore, $A_1 \neq A_2$. Without $A_1 \neq A_2$, Simpson's paradox cannot occur.

Theorem 2. *Simpson's paradox arises only if* $B_1 \neq B_2$.

Proof: Let us assume that $B_1 = B_2$. Then $\beta = bB_1 + B_2(1-b) = bB_1 + B_1(1-b) = B_1 = B_2$. Given that $A_1 > A_2$, it is true that $[aA_1 + A_2(1-a)] > [aA_2 + A_2(1-a)]$. Given that $A_1 > A_2$, it follows that $[aA_1 + A_2(1-a)] > [aA_2 + A_2(1-a)]$. Therefore, $\alpha > A_2$. Yet, $A_2 \geq B_2 = \beta$. So $\alpha > \beta$, which contradicts the assumption. Therefore, $B_1 \neq B_2$. Without $B_1 \neq B_2$, Simpson's paradox cannot occur.

[11] We are indebted to Davin Nelson for the following proofs.