Confirmation, Causation, and Simpson's Paradox

Branden Fitelson[†]

March 12, 2017

1 Simpson's Paradox: The Usual Dialectic

There was a period in the 1970's when the admissions data for the UC–Berkeley graduate school (hereafter, BGS) exhibited some (*prima facie*) peculiar statistical correlations.[1] Specifically, a strong negative correlation was observed between being female and being accepted into BGS. This negative correlation (in the overall population of BGS applicants) was (initially) a cause for some concern regarding the possibility of gender bias in the admissions process at BGS. However, closer scrutiny of the BGS admissions data from this period revealed that *no individual department*'s admissions data exhibited a negative correlation between being female and being admitted. In fact, every department reported a *positive correlation* between being female and being accepted. In other words, a correlation that appears at the level of the general population of BGS applicants is *reversed in every single department* of BGS. This sort of *correlation reversal* is known as *Simpson's Paradox*. Because admissions decisions at BGS are made (autonomously) by each individual department, the lack of departmental correlations seems to rule-out the gender bias hypothesis as the best (causal) explanation of the observed correlations in the data. As it happens, there was a strong positive correlation between being female and applying to a department with a (relatively) high rejection rate. That is, women tended to apply to departments within BGS that have low(er) acceptance rates. And, this is what explains the observed correlations.

This informal presentation is typical of (most) discussions of Simpson's Paradox that one finds in the literature. Typically, one sees *causal explanations* of *how* some set of ("Simpson's-Paradoxical") statistical correlations could have arisen in the first place (*i.e.*, causal "how-possibly" explanations of the "paradoxical" correlational patterns in the statistical data). In this paper, I will be doing something different. I will be discussing what I will call *rationalizing explanations* of the paradoxicality of Simpson's Paradox. That is, I will

[1]I am borrowing this example from Nancy Cartwright [2]. Cartwright was one of the first to offer (probabilistic) causal analyses of Simpson's Paradox. Since then, many statisticians and philosophers have studied the paradox and various causal explanations of how it can arise. See [17] for a nice recent survey.

be examining some explanations of *why* the cases may *seem* paradoxical in the first place.[2] In this sense, my treatment of Simpson's Paradox will be more similar to the way philosophers have approached (many) other "paradoxes" — by trying to *explain (away)* the sense of paradoxicality itself.[3] In the next section, I will provide some background on confirmation theory. In the remaining sections, I will (a) discuss some existing treatments (both causal and confirmation-theoretic) of Simpson's Paradox, and (b) offer a new (purely confirmation-theoretic) rationalizing explanation of the paradoxicality of Simpson's Paradox.

Here is a simplified, precise version of the BGS example, which will serve as our canonical example of Simpson's Paradox for the remainder of the paper. Let there be two departments: E and $\overline{E}$ (think: *easy* and *not* easy) and two genders (of applicants): M and $\overline{M}$ (think: male and *not* male).[4] And, let A ($\overline{A}$) express the proposition than an applicant is *accepted* (*not* accepted). The following table gives all the relevant *acceptance rates* (*i.e.*, statistical frequencies, which we will assume align with the objective probabilities/chances, in this situation) regarding this simplified characterization of BGS admissions (involving a total of 260 applicants, 130 of which applied to department E and 130 of which applied to $\overline{E}$).

	M	$\overline{M}$	Overall
E	$60/80 = \Pr(A \mid E \,\&\, M)$	$40/50 = \Pr(A \mid E \,\&\, \overline{M})$	$100/130 = \Pr(A \mid E)$
$\overline{E}$	$10/50 = \Pr(A \mid \overline{E} \,\&\, M)$	$20/80 = \Pr(A \mid \overline{E} \,\&\, \overline{M})$	$30/130 = \Pr(A \mid \overline{E})$
Overall	$70/130 = \Pr(A \mid M)$	$60/130 = \Pr(A \mid \overline{M})$	$130/260 = \Pr(A)$

Table 1: Probabilistic structure of canonical (BGS) example of Simpson's Paradox

These acceptance rates are to be understood as follows. For instance, $\Pr(A \mid E \,\&\, M) = 60/80$ means that the probability of acceptance for male applicants to department E is $60/80$ — *i.e.*, $60/80$ of the male applicants to department E were accepted (and this frequency accurately reflects the corresponding objective probability of acceptance).

The crucial thing to notice here is that $\overline{M}$ is *negatively* relevant to acceptance in the overall population, *i.e.*, $\Pr(A \mid \overline{M}) < \Pr(A \mid M)$, while $\overline{M}$ is *positively* relevant to acceptance *within each of the two departments*, *i.e.*, $\Pr(A \mid \overline{M} \,\&\, E) > \Pr(A \mid M \,\&\, E)$ and $\Pr(A \mid \overline{M} \,\&\, \overline{E}) >$

---

[2]Such correlational structures are not really paradoxical at all — they occur in nature all the time. I will return to this point later, when I offer my *explanation-away* of the paradoxicality of Simpson's Paradox.

[3]One analogous case that comes to mind here is Hempel's treatment of the Paradox of Confirmation (as explained in [6]). Hempel explains (away) the (apparent) paradoxicality of the paradox of the ravens by (a) isolating an ambiguous presupposition of the "paradox" (which has a false precisficiation and a true precisification), and then (b) arguing that the sense of paradoxicality arises from the conflation of these two conflicting precisifications of the ambiguous presupposition. My strategy for explaining away the paradoxicality of Simpson's Paradox will be very similar (at this level of abstraction) to Hempel's.

[4]The explanations discussed below can all be generalized to cases of Simpson's Paradox that involve m departments and n genders, for any $m, n \geq 2$. But, for simplicity, I will speak only about the 2×2 case.

$\Pr(A \mid M \& \overline{E})$. This kind of *reversal of correlations* (between subpopulations and the overall population) is known as *Simpson's Paradox*.

In the next section, I will provide a bit of theoretical background on confirmation theory. Then, I'll examine two recent discussions of Simpson's Paradox. Finally, I will present a new rationalizing explanation of the (apparent) paradoxicality of Simpson's paradox.

2   Some Background on Confirmation Theory

From a deductive-logical point of view, "argument strength" does not come in degrees. An argument is either deductively valid or it is not. Confirmation theory aims to *generalize* this purely qualitative deductive notion of "argument strength." Ideally, it would be nice to have a *numerical measure* of *the degree to which* some premise P supports some conclusion C. At least, that was Carnap's [1] dream — to construct a *confirmation function* $\mathfrak{c}(C, P)$, which (a) generalizes deductive entailment (in a sense to be made precise, below) and (b) is defined using probabilistic concepts, where (c) the probabilities in terms of which $\mathfrak{c}$ is defined are "objective" (and, ideally, *logical*). For present purposes, we won't need anything like a Carnapian confirmation function. This is for two main reasons. First, we'll only be looking at cases in which there are (uncontroversial, well-defined and well-known) statistical probabilities (*viz.*, *objective chances*). So, we won't need to worry about the "objectivity" of the probability assignments that will underlie our confirmation-theoretic relations (and we won't need to mess around with "logical" probabilities). Second, we'll only need *qualitative* confirmation-theoretic concepts in order to provide our analyses (*i.e.*, numerical confirmation measures won't be required for our purposes). In this sense, we will be employing a more Bayesian (and less Carnapian) approach to confirmation. However, we will follow Carnap in making a key distinction between two kinds of probabilistic confirmation.

> **Confirmation as firmness** (confirms$_f$). P confirms$_f$ C (relative to background supposition S[5]) iff $\Pr(C \mid P \& S) > t$, where t is some (possibly contextual) threshold.

> **Confirmation as increase in firmness** (confirms$_i$). P confirms$_i$ C (relative to background supposition S) iff $\Pr(C \mid P \& S) > \Pr(C \mid S)$.

For present purposes, all of the conditional probabilities $\Pr(\cdot \mid \cdot)$ will be given by a specification of some (uncontroversial) statistical probabilities (*viz.*, objective chances). So, there will

---

[5]When we talk about "background suppositions," we are always thinking of *indicative* suppositions (as opposed to *subjunctive* suppositions). See [11] for more about the concept of "probability relative to an indicative supposition," and how it differs from a *subjunctive*-suppositional conception of $\Pr(\cdot \mid \cdot)$, which may be more appropriate for causal or explanatory applications (*e.g.*, causal analyses of Simpson's Paradox).

be no controversy about where the probabilities in question come from. Our focus here will be on using these two confirmation-theoretic relations to characterize Simpson's Paradox and explain why it *seems* paradoxical (but, in fact, is *not*).

One important theoretical similarity between these two confirmation relations is that they both constitute *generalizations of deductive validity*, in the following sense. If *P entails C*, then *P confirms C, relative to any background supposition S*. To see why, note that if *P* entails *C*, then $\Pr(C \mid P \& S)$ will be *maximal* (*i.e.*, equal to 1), and so we will (generally[6]) have *both* $\Pr(C \mid P \& S) > t$ *and* $\Pr(C \mid P \& S) > \Pr(C \mid S)$. That is, entailing premises will (generally) make their conclusions *highly probable and raise the probabilities* of their conclusions (*above* the values taken relative to whatever background suppositions are involved).

For our present purposes, the key theoretical *differences* between these two notions of confirmation can be seen by thinking of them as constituting (similar, but subtly different) explications of the concept of "argument strength." If a premise *P* confers a high probability on conclusion *C*, then (at least in one sense) the argument can be said to be a "strong" one.[7] However, even if *P* confers a high probability on *C*, it may also be true that *P lowers* the probability of *C* (as compared with its initial/prior probability). In such cases, there also seems to be a sense in which *P* can *not* form the basis of a "strong argument in favor of" *C*. There are many examples of this sort in the literature. Some of the most interesting examples fall under the rubric "the base rate fallacy" [12, 15].[8] For instance, let *P* be the claim that a positive test result has been obtained from an imperfect but reliable test for some rare disease (performed on a sample drawn from some individual *a*). And, let *C* be the claim that *a* does *not* have the rare disease in question. If the disease is sufficiently rare (*i.e.*, if $\Pr(C)$ is sufficiently high), then $\Pr(C \mid P)$ will (still) be very high. Nonetheless, $\Pr(C \mid P)$ will be *less* than $\Pr(C)$, and so *P* will *lower* the probability of *C*. In such a case, it would be odd to say that *P* constitutes strong evidence in favor of *C*. The problem here is that "argument strength" has (at least) two probabilistic components: confirms$_f$ and confirms$_i$, and these two components can *run in opposite directions.* This is what happens in "base rate fallacy" cases (and many other cases besides — see [9] for further discussion). Such cases reveal one important theoretical difference between confirms$_f$ and confirms$_i$.

These two concepts also exhibit some more subtle theoretical divergences that will be im-

---

[6]Strictly speaking, I am assuming here that $\Pr(C \mid S)$ and $\Pr(P \& S)$ do not take extreme values, and that *t* is *strictly between* 1/2 and 1. But, in all the cases we'll be discussing, these (weak) assumptions will hold.

[7]The idea that "argument strength" is best explicated in terms of confirmation as firmness (confirms$_f$) is a very popular one (historically). For instance, Skyrms [19] defends this approach. In general, I think it is a mistake to ignore considerations of increase in firmness (confirms$_i$) when assessing "argument strength." See [9] for a detailed discussion regarding the history of probabilistic explications of "argument strength."

[8]The confirms$_f$/confirms$_i$ distinction is useful for understanding what is going on in *various* "reasoning fallacies" that crop up in the contemporary cognitive psychology literature. See [4, 3] for explanations (away) of the so-called "conjunction fallacy" which make essential use of this distinction.

portant for our analysis of Simpson's Paradox. In order to grok these more subtle theoretical differences, it is useful to consider the following principle regarding "argument strength."

> **The *Suppositional* Sure Thing Principle** (SSTP).[9] If P constitutes a strong argument for C *on the supposition that S* and P constitutes a strong argument for C *on the supposition that $\overline{S}$*, then P constitutes a strong argument for C (*unconditionally*).

Basically, the SSTP involves a sort of suppositional "reasoning by cases." It asks us to consider whether (a) the argument $P \therefore C$ is strong, given some (indicative) supposition S and (b) the argument $P \therefore C$ is strong, given *the negation of* that (indicative) supposition (*viz.*, $\overline{S}$). And, if both of these assessments come out positive, then the SSTP implies that $P \therefore C$ is strong *unconditionally* (*i.e.*, in the *absence* of background suppositions).

This kind of suppositional "reasoning by cases" is correct for some understandings of "argument strength" but not for others. Specifically, SSTP is satisfied by confirmation as firmness (confirms$_f$), but *not* by confirmation as increase in firmness (confirms$_i$). That is to say, we have the following two key theoretical facts.[10]

> **Fact 1**. Confirmation$_f$ *satisfies* SSTP. In other words, it is generally true that if (a) P confirms$_f$ C, relative to S and (b) P confirms$_f$ C, relative to $\overline{S}$, then (c) P confirms$_f$ C, unconditionally.

> **Fact 2**. Confirmation$_i$ *violates* SSTP. In other words, there exist cases/contexts in which (a) P confirms$_i$ C, relative to S and (b) P confirms$_i$ C, relative to $\overline{S}$; *but*, (c) P does *not* confirm$_i$ C, unconditionally.

Fact 2 is the theoretical property of confirms$_i$ which makes Simpson's Paradox possible in the first place. Indeed, Simpson's Paradox cases *just are* the examples which establish the truth of Fact 2 (*e.g.*, the simplified BGS example depicted in Table 1 above).

While confirms$_f$ and confirms$_i$ *disagree* regarding the SSTP, they *agree* on a similar (but subtly different) theoretical principle regarding "argument strength." To wit:

> **The *Conjunctive* Sure Thing Principle** (CSTP). If $P \& S$ constitutes a strong argument for C (*unconditionally*) and $P \& \overline{S}$ constitutes a strong argument for C (*unconditionally*), then P constitutes a strong argument for C (*unconditionally*).

---

[9]The "sure thing" moniker is borrowed from Savage [18], who uses it to describe a similar "reasoning by cases" property of preference relations.

[10]See the APPENDIX for elementary proofs of these two key Facts, as well as Fact 3 below.

Like the SSTP, the CSTP also involves "reasoning by cases." However, unlike the SSTP (which involves *suppositional* reasoning), the CSTP involves *conjunctive* (and *non*-suppositional) "reasoning by cases." That is, the CSTP does *not* ask us to assess the strength of a *single* argument $P \therefore C$, relative to a *pair* of *suppositions* (*viz.*, S and $\overline{S}$). Rather, the CSTP asks us to consider the *unconditional* strength of a *pair* of arguments ($P \& S \therefore C$ and $P \& \overline{S} \therefore C$). Interestingly, *both* confirmation$_f$ *and* confirmation$_i$ *satisfy* the CSTP. That is, we have:

> **Fact 3**. Confirmation$_f$ *and* confirmation$_i$ *satisfy* CSTP. In other words, it is generally true that if (a) $P \& S$ confirms C and (b) $P \& \overline{S}$ confirms C, then (c) P confirms C.[11]

The disagreement between confirms$_f$ and confirms$_i$ with respect to SSTP *vs.* CSTP can be understood in terms of their *ability to distinguish* conjunctive *vs.* suppositional support/confirmation. Confirmation as firmness simply *cannot* distinguish between the two. If you want to know whether P confirms$_f$ C, on the supposition that S, you check whether $\Pr(C \mid P \& S) > t$. And, this is *exactly the same thing* you must do if you want to check whether $P \& S$ confirms$_f$ C, unconditionally. In other words, from the point of view of confirmation as firmness, *there is no distinction* between conjunctive and suppositional support (in general). In light of this, it is easy to see why Facts 1 and 3 both obtain (for confirms$_f$) — they are actually *one and the same fact*, from the perspective of confirmation$_f$. This is not so for confirmation$_i$, which distinguishes suppositional and conjunctive support in such a way that allows for Simpson's-Paradoxical cases (*i.e.*, violations of the SSTP) to arise.

Facts 1–3 are the key theoretical facts that will undergird my (ultimate) rationalizing explanation of Simpson's Paradox. In the next section, I will examine two recent discussions of Simpson's Paradox. This will lead naturally into a new rationalizing explanation, based on Facts 1–3 above.

3 Two Recent Discussions of Simpson's Paradox

It is clear that there can be nothing truly *paradoxical* about "Simpson's Paradox." After all, examples like Cartwright's BGS example *actually occur in nature* (all the time). So, what needs to be explained (or perhaps explained *away*) is the *appearance* of paradoxicality in cases of Simpson's Paradox (*i.e.*, their *prima facie* paradoxicality). In this section, I'll discuss two recent treatments of Simpson's Paradox. Then, I'll propose a new, purely confirmation theoretic rationalizing explanation.

---

[11]Unsurprisingly, it's also true that *deductive entailment* satisfies CSTP. That is, if $P \& S \vDash C$ and $P \& \overline{S} \vDash C$, then $P \vDash C$. Thus, on pretty much any way of explicating "argument strength," the CSTP will come out true.

3.1 Pearl's Causal Rationalizing Explanation

Pearl [16, §6.1] provides a *causal* rationalizing explanation of Simpson's Paradox, according to which its apparent paradoxicality involves the conflation of an invalid argument involving confirmation$_i$ and a valid argument involving *causation*. Let's use our BGS example to illustrate Pearl's approach. In that example, there is a correlation between being female and being rejected from graduate school. But, this correlation is *spurious*. It is explained by the fact that females tend to apply to departments with lower acceptance rates. Furthermore, this "tendency" is *not* itself *causal*. How can we tell? Pearl would explain this by postulating that if we were to *intervene* on an applicant's gender, this would not have an effect on the probability that they apply to one department rather than another. In Pearl's notation, this amounts to the following assumption:

(†) $\Pr(E \mid do(M)) = \Pr(E \mid do(\overline{M})) = \Pr(E).$[12]

The main idea behind Pearl's rationalizing explanation is that people (to the extent they see the BGS case as *paradoxical*) are (falsely) assuming that the correlations involved all reflect *causal* relationships. In Pearl's notation, we can express the three key correlations in the BGS case as *causal* relations, in the following way:

(1) $\Pr(A \mid do(M) \,\&\, E) < \Pr(A \mid do(\overline{M}) \,\&\, E)$

(2) $\Pr(A \mid do(M) \,\&\, \overline{E}) < \Pr(A \mid do(\overline{M}) \,\&\, \overline{E})$

(3) $\Pr(A \mid do(M)) > \Pr(A \mid do(\overline{M}))$

Claim (1) asserts that being male *causally decreases* the chance of acceptance — within department E. Claim (2) asserts that being male *causally decreases* the chance of acceptance — within department $\overline{E}$. And, claim (3) asserts that being male *causally increases* the chance of acceptance — in the overall population.

It can be shown (1)–(3) *cannot all be true*, if gender has no causal influence on which department one applies to (*i.e.*, if claim (†) is true). That is, assuming (†), it is impossible for the three key (Simpson-reversal) correlations in the BGS case to reflect *causal* relationships. And, according to Pearl, this explains the sense of paradoxicality one may feel when one first hears about (*e.g.*) the BGS case. For *if* all of the correlations involved in the reversal

---

[12]Here, "$do(X)$" means "the value of X is set by an *intervention* on X." Where, Pearl-style interventions are ways of setting the value of X, which cancel any other causal influences that may be upstream from X. Thus, interventions are ways of revealing whether there is really is a direct causal connection between X and some potential (downstream) effect of X. Here, the truth of (†) would imply that the "tendency" of females to apply to more selective departments is *not* a *causal* relationship.

were causal, *then* — assuming (†) — this *would* be paradoxical (as it would involve an *inconsistent* set of probabilistic claims). Thus, Pearl's account predicts that once one learns the underlying causal structure of the case (*e.g.*, that condition (†) holds in the case, *etc.*), one's sense of paradoxicality should disappear.[13]

Pearl's causal rationalizing explanation is interesting and illuminating. But, I think it has a couple of potential shortcomings. First, it is not a purely confirmation-theoretic explanation. Simpson's Paradox does not essentially involve any causal concepts. It only (essentially) involves relations of confirmation$_i$. As such, it would be nice to have a purely confirmation-theoretic rationalizing explanation. Second, while Pearl's account does involve the conflation of a valid argument (his causal argument) and an invalid one (the purely confirmation$_i$-theoretic one), the valid argument is *enthymematic*. That is, Pearl's causal paradox does not just involve the correlations that are given in the statement of the paradox [*i.e.*, (1)–(3)] it also involves an *unstated causal independence* premise/assumption [*i.e.*, (†)] which is not essential to the statement of Simpson's Paradox. In these two senses, Pearl's rationalizing explanation brings in considerations that are not essential to the Paradox (*per se*). It would be preferable (*ceteris paribus*) to have a rationalizing explanation that appeals only to information that is given (essentially) in the statement of the paradoxical cases.

Next, I will discuss Kotzen's [14] recent take on Simpson's Paradox. That will lead us toward our purely confirmation-theoretic rationalizing explanation.

3.2   Kotzen on Simpson's Paradox and The Miners Puzzle

Kotzen [14] draws an analogy between Simpson's Paradox and the Miners Puzzle [13]. For our purposes, the key feature of Kotzen's approach is the reconstruction of Simpson's paradoxical cases as violations of the following classically valid form of deductive inference:

If S, then P confirms$_i$ Q.

If $\overline{S}$, then P confirms$_i$ Q.

Therefore, P confirms$_i$ Q.

---

[13]Miriam Schoenfield (personal communication) has conjectured a similar, causal rationalizing explanation. Hers is expressed in terms of subjunctive conditional probabilities [11] rather than Pearl's [16] *do*-calculus. But, like Pearl, Schoenfield will need to assume that the causal independence expressed by (†) holds in her examples. Schoenfield has done some preliminary experiments which indicate a decrease in the sense of paradoxicality of Simpson's Paradox cases, when one learns about the causal structure of the case. It would be interesting to do some experiments which also involve my suppositional/conjunctive distinction [and which involve possible violations of (†)], so as to tease apart the roles of causal and non-causal influences on our sense of paradoxicality in cases of Simpson's Paradox. This is fertile ground for work in experimental philosophy (and the psychology of inductive/causal inference). See, also, *fn.* 17, below.

While this form of argument is valid for the *material* conditional (*i.e.*, if $\ulcorner$if p, then $q\urcorner$ is interpreted as $\ulcorner p \supset q \urcorner$), it may not be valid for other kinds of conditionals (*e.g.*, the English indicative conditional). Some authors have taken the Miners Puzzle to provide reasons to think that the English indicative does not obey various classically valid forms of deductive inference (*e.g.*, the form of inference above, or even *modus ponens*), when the consequent(s) of the conditionals contain various sorts of *modal* expressions [13]. Similarly, Kotzen presents Simpson's Paradox as having an analogous structure/upshot — but with respect to consequent(s) containing *confirmation$_i$-theoretic* expressions.

While Kotzen's aim is not to provide a rationalizing explanation of why Simpson's Paradox seems paradoxical in the first place, I think his analogy with the Miners Puzzle puts us on the right track toward formulating such an explanation.

4   A New Rationalizing Explanation of Simpson's Paradox

I think Kotzen's discussion leads us in the direction of a *purely confirmation-theoretic* rationalizing explanation of Simpson's Paradox. I propose that the *prima facie* paradoxicality of Simpson's Paradox can be explained (at least in part) by appealing to the (subtle) distinction between *conjunctive* confirmation (or support) and *suppositional* confirmation (or support) that was implicit in our discussion of SSTP *vs.* CSTP above. To wit, contrast the following:

> **Suppositional Reading**. If S, then P confirms C.

> **Conjunctive Reading**. $P \mathbin{\&} S$ confirms C.[14]

I propose that there is a tendency for hearers of descriptions of Simpson's Paradox cases to *conflate* these two readings. Let's apply this to our BGS case. To make things more concrete, let us now suppose that department $\overline{E}$ is History. Then, we may describe the correlation between being female and being accepted to History, in the following two ways:

> **Suppositional**. If one applies to History, then being female improves one's chances of acceptance.

> **Conjunctive**. Being a female History applicant improves one's chances of acceptance.

---

[14]Readers familiar with controversies regarding the *import-export* principle for the indicative conditional [7, 10] will notice that there are some structural similarities between these two controversies. In fact, if one adopts a semantics for the indicative conditional that involves a probabilistic relevance/confirmation$_i$ component [5], then the two controversies become (almost) *one and the same*. This is important, since many people think that import-export is *true* (for the English indicative), which suggests that people may well be susceptible to the kind of conflation I exploit in my rationalizing explanation, below.

In our BGS example, it is *only* the **Suppositional** reading that is true.[15] Moreover, as we have seen, these two readings — while similar-sounding to the (untrained) ear — make all the difference, in terms of *paradoxicality*. If we slide from **Suppositional** to **Conjunctive**, then we are sliding from an interpretation on which the examples described are *possible* (because of Fact 2), to one on which they are *impossible* (because of Fact 3). And, this will be true for *any* example of Simpson's Paradox — even ones in which Pearl's causal independence assumption (†) *fails*. So, if my proposal is right, then we can exploit the fact that the *conjunctive* STP is a *truism* about confirmation (no matter how it is explicated) to furnish an explanation-*away* of the *apparent* paradoxicality of Simpson's Paradox.

Here's how that explanation would go. Examples of Simpson's Paradox involve violations of the SSTP. To the extent that hearers tend to conflate suppositional *vs.* conjunctive readings of the key locutions involved in descriptions of such cases, there will be a tendency to hear such cases as violations of the **C**STP. But, *there can be no violations of the CSTP* (since, as Fact 3 reveals, it is a general truism about confirmation — no matter how it is explicated). Therefore, to the extent that hearers tend to conflate suppositional *vs.* conjunctive readings of the key locutions involved in descriptions of such cases, there will be a tendency for hearers to interpret the cases as implying something *paradoxical* (*i.e.*, as somehow *contradicting a theoretical truism*).[16] But, because there can be no such cases, the appearance of paradoxicality is a *mere* appearance (arising, at least in part, from said conflation).

Our rationalizing explanation has a couple of virtues, in comparison with Pearl's causal rationalizing explanation. First, ours is *purely confirmation-theoretic* — it makes no appeal to causal relations, which are *inessential* to the description of Simpson's Paradoxical cases. Moreover, Pearl's explanation relies on an *implicit causal independence assumption* [(†)] which is not stated in the descriptions of the Paradox. Our explanation applies to *all* instances of Simpson's Paradox — even ones in which Pearl's assumption (†) *fails*.[17] In any case, even if causal facts/presuppositions are playing a role in the (*prima facie*) paradoxicality of Simpson's Paradox, my proposal is that the suppositional/conjunctive distinc-

---

[15]In our BGS example, being a female history applicant actually *lowers* one's probability of acceptance, since $\Pr(A \mid \overline{E} \, \& \, \overline{M}) = 1/4 < 1/2 = \Pr(A)$. So, the **Conjunctive** reading is *false* in the BGS example.

[16]One reason for this conflation might be people's (rather general) tendency to conflate confirms$_f$ and confirms$_i$ (recall, *only* confirms$_i$ *is capable of distinguishing* conjunctive and suppositional confirmation). This sort of conflation has been seen in other contexts as well. See [9, 4, 3] for further discussion. Another possible reason for the conflation is the general acceptance of the truth of import-export for the indicative conditional [10]. See, also, *fn.* 14, above.

[17]It is important to note that Pearl's assumption (†) is *very strong*. It requires that $\Pr(E \mid do(M)) = \Pr(E \mid do(\overline{M})) = \Pr(E)$. I suspect there will be cases in which *either* $\Pr(E \mid do(M)) \neq \Pr(E \mid do(\overline{M}))$, *or* $\Pr(E \mid do(\overline{M})) \neq \Pr(E)$, *or* $\Pr(E \mid do(M)) \neq \Pr(E)$; but, some (residual) sense of paradoxicality is *nonetheless* felt in hearers of the (non-causal description of) the case. Pearl's explanation (and other similar causal explanations) *goes silent* in *all* such cases, whereas ours still applies (equally well). These sorts of empirical questions would be interesting to explore, experimentally. See, also, *fn.* 13 above.

tion/conflation also has (independent, rationalizing) explanatory relevance/force.

APPENDIX: Proofs of Facts 1–3

For our proofs, we will adopt the algebraic style/methods of [8]. The following *stochastic truth table* [8] represents all possible probability distributions over the $\{P, C, S\}$ language, *via* the 7 real variables $a, b, c, d, e, f, g \in [0, 1]$.

P	C	S	$\Pr(\cdot)$
T	**T**	**T**	a
T	**T**	**F**	b
T	**F**	**T**	c
T	**F**	**F**	d
F	**T**	**T**	e
F	**T**	**F**	f
F	**F**	**T**	g
F	**F**	**F**	$1 - (a + b + c + d + e + f + g)$

Fact 1. Confirmation$_f$ *satisfies* SSTP. In other words, it is generally true that if (a) P confirms$_f$ C, relative to S and (b) P confirms$_f$ C, relative to $\overline{S}$, then (c) P confirms$_f$ C, unconditionally.

Proof. Fact 1 asserts that the following argument in probability calculus is valid (where t is some positive threshold value on the unit interval).

$$(\mathcal{A}_1) \quad \frac{\begin{array}{l} (1)\ \Pr(C \mid P \,\&\, S) > t. \\ (2)\ \Pr(C \mid P \,\&\, \overline{S}) > t. \end{array}}{\therefore (3)\ \Pr(C \mid P) > t.}$$

Algebraically (using the above setup [8]), $\mathcal{A}_1$ is equivalent to:

$$(\mathcal{A}_1) \quad \frac{\begin{array}{l} (1)\ \frac{a}{a+c} > t. \\ (2)\ \frac{b}{b+d} > t. \end{array}}{\therefore (3)\ \frac{a+b}{a+b+c+d} > t.}$$

Cross-multiplying (1) & (2) yields:

(1) $\qquad\qquad\qquad\qquad a > (a + c) \cdot t$

(2) $\qquad\qquad\qquad\qquad b > (b + d) \cdot t$

11

Adding the lhs & rhs of these and collecting t yields $\quad \square$

$$a + b > (a + b + c + d) \cdot t$$

which is equivalent to (3). $\hfill \square$

Fact 2. Confirmation$_i$ *violates* SSTP. In other words, there exist cases/contexts in which (a) P confirms$_i$ C, relative to S and (b) P confirms$_i$ C, relative to $\overline{S}$; *but*, (c) P does *not* confirm$_i$ C, unconditionally.

Proof. The BGS example depicted in Table 1 is sufficient to establish this claim. For completeness, I include here the full stochastic truth table representation of the probability model depicted in Table 1 above (in this example, the relevant instantiation of the SSTP schema is: $P \stackrel{\text{def}}{=} \overline{M}$, $S \stackrel{\text{def}}{=} E$, and $C \stackrel{\text{def}}{=} A$).

A	E	M	$\Pr(\cdot)$
T	**T**	**T**	$3/13$
T	**T**	**F**	$2/13$
T	**F**	**T**	$1/26$
T	**F**	**F**	$1/13$
F	**T**	**T**	$1/13$
F	**T**	**F**	$1/26$
F	**F**	**T**	$2/13$
F	**F**	**F**	$3/13$

It is easily verified that, in this model, the following three constraints are satisfied:

(i) $\overline{M}$ confirms$_i$ A, relative to E.

$$\Pr(A \mid \overline{M} \,\&\, E) = 4/5 > 3/4 = \Pr(A \mid M \,\&\, E)$$

(ii) $\overline{M}$ confirms$_i$ A, relative to $\overline{E}$.

$$\Pr(A \mid \overline{M} \,\&\, \overline{E}) = 1/4 > 1/5 = \Pr(A \mid M \,\&\, \overline{E})$$

(ii) $\overline{M}$ *dis*confirms$_i$ A, *unconditionally*.

$$\Pr(A \mid \overline{M}) = 6/13 < 7/13 = \Pr(A \mid M)$$

12

Thus, the BGS example illustrates the failure of confirms$_i$ to satisfy the SSTP. $\square$

Fact 3. Confirmation$_f$ *and* confirmation$_i$ *satisfy* CSTP. In other words, it is generally true that if (a) $P \& S$ confirms C and (b) $P \& \overline{S}$ confirms C, then (c) P confirms C.

Proof. We have already shown that confirmation$_f$ satisfies CSTP, since this is equivalent to showing that argument form $\mathcal{A}_1$, above, is valid.

To show that confirmation$_i$ satisfies CSTP, we just need to show that the following argument form is valid.

$$(\mathcal{A}_2) \quad \frac{\begin{array}{l} (1)\ \Pr(C \mid P \& S) > \Pr(C). \\ (2)\ \Pr(C \mid P \& \overline{S}) > \Pr(C). \end{array}}{\therefore (3)\ \Pr(C \mid P) > \Pr(C).}$$

Algebraically (using the above setup [8]), $\mathcal{A}_2$ is equivalent to

$$(\mathcal{A}_2) \quad \frac{\begin{array}{l} (1)\ \frac{a}{a+c} > a + b + e + f. \\[4pt] (2)\ \frac{b}{b+d} > a + b + e + f. \end{array}}{\therefore (3)\ \frac{a+b}{a+b+c+d} > a + b + e + f.}$$

Cross-multiplying (1) & (2) yields

$$(1) \qquad\qquad\qquad a > (a + c) \cdot (a + b + e + f)$$

$$(2) \qquad\qquad\qquad b > (b + d) \cdot (a + b + e + f)$$

Adding the lhs & rhs of these and collecting t yields

$$a + b > (a + b + c + d) \cdot (a + b + e + f)$$

which is equivalent to (3). $\square$

References

[1] R. Carnap. *Logical foundations of probability*, 2nd ed. University of Chicago Press, 1962.

[2] N. Cartwright. "Causal Laws and Effective Strategies," *Noûs* (1979): 419-437.

[3] V. Crupi, K. Tentori and S. Russo. "On the determinants of the conjunction fallacy: Probability vs. inductive confirmation," *Journal of Experimental Psychology: General* (2013): 235–255.

[4] V. Crupi, K. Tentori and B. Fitelson. "Probability, Confirmation and the Conjunction Fallacy," *Thinking & Reasoning* (2008): 182–199.

[5] I. Douven, "The Evidential Support Theory of Conditionals," *Synthese* (2009): 19–44.

[6] B. Fitelson and J. Hawthorne. "The Wason Task(s) and the Paradox of Confirmation," *Philosophical Perspectives* (2010): 207–241.

[7] B. Fitelson. "The Strongest Possible Lewisian Triviality Result," *Thought* (2015): 69–74.

[8] _____, "A decision procedure for probability calculus with applications," Review of Symbolic Logic (2008): 111–125.

[9] _____, *Inductive Logic*, in *The Philosophy of Science: An Encyclopedia*, J. Pfeifer and S. Sarkar (*eds.*), Routledge, 2005.

[10] A. Gillies, "Indicative Conditionals," in D. Graff Fara and G. Russell (*eds.*), The Routledge Companion to Philosophy of Language, Routledge, 2012: 449–465.

[11] J. Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, 1999.

[12] D. Kahneman and A. Tversky. "Evidential impact of base rates," in D. Kahneman, P. Slovic and A. Tversky (*eds.*), *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, 1982: 153–160.

[13] N. Kolodny and J. MacFarlane, "Ifs and Oughts," *Journal of Philosophy* (2010): 115-143.

[14] M. Kotzen, "Conditional Oughts and Simpson's Paradox," unpublished manuscript, 2013.

[15] I. Levi, "Who commits the base rate fallacy?" *Behavioral and Brain Sciences* (1983): 502–506.

[16] J. Pearl, *Causality: Models, Reasoning, and Inference*, Second Edition, Cambridge University Press, 2009.

[17] G. Malinas and J. Bigelow, "Simpson's Paradox," *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), E. Zalta (*ed.*), URL =
`<http://plato.stanford.edu/archives/sum2016/entries/paradox-simpson/>`.

[18] L. Savage, *The Foundations of Statistics*, Dover, 1972.

[19] B. Skyrms, *Choice and Chance: An Introduction to Inductive Logic*, 4th ed. Wadsworth, 2000.