

Graduate School Admissions (GSA).¹ A graduate school became suspicious when it noticed a *negative correlation* between being female and being accepted (across *all* applicants in a given year). Further investigation revealed that — *within each department* — there was a *positive correlation* between being female and being accepted!

- Assume *two* departments: E and \bar{E} (think: *Easy & not-Easy*). Let $M \stackrel{\text{def}}{=} \text{male}$, $\bar{M} \stackrel{\text{def}}{=} \text{female}$ (gender binary), $A \stackrel{\text{def}}{=} \text{accepted}$, and $\bar{A} \stackrel{\text{def}}{=} \text{rejected}$. Finally, assume these *acceptance rates*.

	M	\bar{M}	Overall
E	60/80	40/50	100/130
\bar{E}	10/50	20/80	30/130
Overall	70/130	60/130	130/260

¹This is a simplified version [10] of the example originally discussed in [2]. See vudlab.com/simpsons/ for some fun visualizations of this example.

- I'll offer a "rationalizing explanation" of *why* such examples may *seem* "paradoxical" shortly. First, I'll say some things about the existing literature on Simpson's Paradox (SP).
- The existing literature on SP tends to be concerned (almost exclusively) with doing the following two kinds of things.

- Giving *mathematical explanations* (or *characterizations*) of the (algebraic) *structure* of SP cases. The 2×2 case (as above), has a simple algebraic characterization [10].²
- Giving *causal explanations* of the statistical distributions observed in *actual* SP cases [11]. In GSA, the causal story is simple (and benign). Females tend to apply to the "harder" department, *i.e.*, the department with lower acceptance rate.

- My aim is different — I want a "rationalizing explanation" of *why* SP may (reasonably) *seem paradoxical* in the first place.

²The probability calculus (a decidable algebraic theory [8]) can be used to provide a *general* characterization of the 2×2 case. See Extras slides 11–12.

- There are various ways to try to explain (or *explain-away*) "paradoxicality." My approach fits the following mold [5].
 - Step 1.** Identify the (or *an* explanatorily salient) argument form \mathcal{A} that is *invalidated* by the "paradoxical" cases.
 - Step 2.** Identify a "similar" or "nearby" (more on this below) form of argument form \mathcal{A}^* , which is *universally valid*.
 - Step 3.** Argue (or invite the listener to consider) that the (reasonable) *appearance* of paradoxicality stems (at least, in part) from *conflating* \mathcal{A} and \mathcal{A}^* — that is, from hearing examples of the "paradox" as invalidating argument form \mathcal{A}^* (which *would be* paradoxical, since \mathcal{A}^* is *valid*).
- Ideally, the argument forms \mathcal{A} and \mathcal{A}^* should:
 - have sufficiently similar *logical forms*, and
 - employ sufficiently similar (probabilistic) *concepts*.
- Moreover, ideally, the "explanation" should (c) be *fully general* (*i.e.*, be applicable to *all instances* of the "paradox").

- Carnap [1] distinguished two senses of "confirmation."
 - Confirmation as firmness.** P confirms _{f} Q , on the (indicative) supposition that R , just in case Q is (sufficiently) *probable*, conditional upon the conjunction $P \& R$. Formally,

$$P \text{ confirms}_f Q, \text{ on the (indicative) supposition that } R$$

iff

$$\Pr(Q | P \& R) > t.^3$$
 - Confirmation as increase in firmness.** P confirms _{i} Q , on the (indicative) supposition that R , just in case P and Q are *positively correlated*, conditional upon R . Formally,

$$P \text{ confirms}_i Q, \text{ on the (indicative) supposition that } R$$

iff

$$\Pr(Q | P \& R) > \Pr(Q | \bar{P} \& R).$$

- Since our "explanation" of SP will be formal/universal, the *interpretation* of $\Pr(\cdot | \cdot)$ can be allowed to vary, as needed.

³Here, $t \geq 1/2$ is some (possibly contextually determined) *threshold*.

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○●○○○○○	Extras ○○○○○	References
<ul style="list-style-type: none"> ● That brings us to Step 1 of our "explanation." Here is the argument form \mathcal{A} that is invalidated by instances of SP.⁴ <div style="margin-left: 20px;"> <p>(1) P confirms_{i} Q, on the supposition that R.</p> <p>(\mathcal{A}) (2) P confirms_{i} Q, on the supposition that \bar{R}.</p> <hr style="width: 80%; margin-left: 0;"/> <p>∴ (3) P confirms_{i} Q, <i>unconditionally</i>.⁵</p> </div> ● In our GSA example, \mathcal{A} is instantiated as follows: <div style="margin-left: 20px;"> <p>(1) \bar{M} confirms_{i} A, on the supposition that E.</p> <p>(2) \bar{M} confirms_{i} A, on the supposition that \bar{E}.</p> <hr style="width: 80%; margin-left: 0;"/> <p>∴ (3) \bar{M} confirms_{i} A, <i>unconditionally</i>.</p> </div> ● The Pr-distribution determined by our GSA 2×2 table above constitutes a <u>counterexample</u> to the validity of \mathcal{A}. <p style="margin-left: 20px;">⁴Strictly speaking, SP is more general than \mathcal{A}, since (i) it can also involve <i>disconfirmation_{i}</i> and/or <i>irrelevance</i>, and (ii) it can involve random variables with <i>more than two values</i>. My explanation(s) go through in full generality.</p> <p style="margin-left: 20px;">⁵Note: "unconditionally" just means "on a tautological supposition."</p> 			
Branden Fitelson	Confirmation, Causation, and Simpson's Paradox	5	

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○●○○○○	Extras ○○○○○	References																
<ul style="list-style-type: none"> ● To see this, we can add <i>probabilistic labels</i> to our GSA table. <table border="1" style="margin: 10px auto; border-collapse: collapse; text-align: center;"> <thead> <tr style="border-top: 2px solid black; border-bottom: 2px solid black;"> <th style="border: none;"></th> <th style="border: none;">M</th> <th style="border: none;">\bar{M}</th> <th style="border: none;">Overall</th> </tr> </thead> <tbody> <tr> <td style="border: none;">E</td> <td style="border: none;">$60/80 = \Pr(A E \& M)$</td> <td style="border: none;">$40/50 = \Pr(A E \& \bar{M})$</td> <td style="border: none;">$100/130 = \Pr(A E)$</td> </tr> <tr> <td style="border: none;">\bar{E}</td> <td style="border: none;">$10/50 = \Pr(A \bar{E} \& M)$</td> <td style="border: none;">$20/80 = \Pr(A \bar{E} \& \bar{M})$</td> <td style="border: none;">$30/130 = \Pr(A \bar{E})$</td> </tr> <tr style="border-bottom: 2px solid black;"> <td style="border: none;">Overall</td> <td style="border: none;">$70/130 = \Pr(A M)$</td> <td style="border: none;">$60/130 = \Pr(A \bar{M})$</td> <td style="border: none;">$130/260 = \Pr(A)$</td> </tr> </tbody> </table> ● With these probabilistic labels in place, we can now see that: <ol style="list-style-type: none"> (1) $\Pr(A E \& M) < \Pr(A E \& \bar{M})$. That is, \bar{M} confirms_{i} A, on the supposition that E. In words: being female is <i>positively correlated</i> with acceptance, <i>in department E</i>. (2) $\Pr(A \bar{E} \& M) < \Pr(A \bar{E} \& \bar{M})$. That is, \bar{M} confirms_{i} A, on the supposition that \bar{E}. In words: being female is <i>positively correlated</i> with acceptance, <i>in department E</i>. (3) $\Pr(A M) > \Pr(A \bar{M})$. That is, \bar{M} <i>disconfirms_{i}</i> A, <i>unconditionally</i>. In words: being female is <i>negatively correlated</i> with acceptance, <i>in the general population</i>. ● This explains, purely in confirmation_{i} terms, what SP is. 					M	\bar{M}	Overall	E	$60/80 = \Pr(A E \& M)$	$40/50 = \Pr(A E \& \bar{M})$	$100/130 = \Pr(A E)$	\bar{E}	$10/50 = \Pr(A \bar{E} \& M)$	$20/80 = \Pr(A \bar{E} \& \bar{M})$	$30/130 = \Pr(A \bar{E})$	Overall	$70/130 = \Pr(A M)$	$60/130 = \Pr(A \bar{M})$	$130/260 = \Pr(A)$
	M	\bar{M}	Overall																
E	$60/80 = \Pr(A E \& M)$	$40/50 = \Pr(A E \& \bar{M})$	$100/130 = \Pr(A E)$																
\bar{E}	$10/50 = \Pr(A \bar{E} \& M)$	$20/80 = \Pr(A \bar{E} \& \bar{M})$	$30/130 = \Pr(A \bar{E})$																
Overall	$70/130 = \Pr(A M)$	$60/130 = \Pr(A \bar{M})$	$130/260 = \Pr(A)$																
Branden Fitelson	Confirmation, Causation, and Simpson's Paradox	6																	

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○●○○○○	Extras ○○○○○	References
<ul style="list-style-type: none"> ● So much for Step 1. On to Step 2. I will examine <i>two</i> confirmation-theoretic (and 1 causal) "explanations" of SP. ● First Way. This way involves a postulated simple conflation of Carnap's two senses of confirmation (note: this is a well-established human psychological conflation [3]). <div style="margin-left: 20px;"> <p>(1) P confirms_{f} Q, on the supposition that R.</p> <p>(\mathcal{A}_1^*) (2) P confirms_{f} Q, on the supposition that \bar{R}.</p> <hr style="width: 80%; margin-left: 0;"/> <p>∴ (3) P confirms_{f} Q, <i>unconditionally</i>.</p> </div> ● \mathcal{A}_1^* is <i>universally valid</i>. This is easy to show, since it (essentially) boils down to the following "most" validity.⁶ <div style="margin-left: 20px;"> <p>(1) Most P & R-worlds are Q-worlds.</p> <p>(\mathcal{A}_1^*) (2) Most P & \bar{R}-worlds are Q-worlds.</p> <hr style="width: 80%; margin-left: 0;"/> <p>∴ (3) Most P-worlds are Q-worlds.</p> </div> <p style="margin-left: 20px;">⁶See Extras slide 13 for a rigorous algebraic proof of the validity of \mathcal{A}_1^*.</p> 			
Branden Fitelson	Confirmation, Causation, and Simpson's Paradox	7	

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○●○○○○	Extras ○○○○○	References
<ul style="list-style-type: none"> ● Second Way. This way does not involve a simple conflation of Carnap's two senses of "confirms." It involves <i>only</i> confirmation as increase in firmness (confirms_{i}). ● This time, the conflation will involve what I will call <i>suppositional vs conjunctive</i> confirmation. ● Here's an analogy to help get a grip on the distinction. Consider the following two indicative conditional forms: <ol style="list-style-type: none"> (I) If R, then if P then Q. (II) If P & R, then Q. ● Many philosophers [9] have claimed that (I) and (II) are (in general) <i>equivalent</i>.⁷ As such, many think conflating (I) and (II) is OK. Here's an analogous <i>confirmational</i> pair. <ol style="list-style-type: none"> (S) P confirms Q, on the supposition that R. (C) P & R confirms Q, <i>unconditionally</i>. <p style="margin-left: 20px;">⁷This equivalence is called <i>import-export</i>. See [7, 4, 6] for discussion.</p> 			
Branden Fitelson	Confirmation, Causation, and Simpson's Paradox	8	

Example of SP ○ Confirmation-Theoretic "Explanation" of SP ○○○○○○● Extras ○○○○ References

- Interestingly, \mathcal{S} and \mathcal{C} are *equivalent* for confirms_f , but *not* for confirms_i . To see the former, simply apply the definition of confirms_f above (then the equivalence will be obvious).
- The fact that \mathcal{S} and \mathcal{C} are *not* equivalent for confirms_i is the key to our Second Way. To wit, here's my second \mathcal{A}^* form.
 - (1) $P \ \& \ R$ confirms Q , *unconditionally*.
 - (2) $P \ \& \ \bar{R}$ confirms Q , *unconditionally*. \therefore (3) P confirms Q , *unconditionally*.

☞ \mathcal{A}_2^* is *valid* for *both* confirms_f and confirms_i .⁸

- So, if (3) is false, then at least one of (1) and (2) must also be false. In our GSA example, we have the following contrast:
 - (S) \bar{M} confirms_{*i*} A , *on the supposition that* \bar{E} .
 - (C) $\bar{M} \ \& \ \bar{E}$ *does not* confirm_{*i*} A , *unconditionally*.

⁸See Extras slides 13 and 14 for algebraic proofs of these claims.

Branden Fitelson Confirmation, Causation, and Simpson's Paradox 9

Example of SP ○ Confirmation-Theoretic "Explanation" of SP ○○○○○○● Extras ○○○○ References

- How do our Two Ways fare wrt our three *desiderata* (a)–(c)?
- \mathcal{A}_1^* (a) has *the same* logical form as \mathcal{A} , (b) involves *two* confirmational concepts (but those concepts are conflated in other contexts [3, 4]), and (c) covers *all* SP cases.
- \mathcal{A}_2^* (a) has a *different* logical form than \mathcal{A} (but the two forms are conflated in other contexts [9, 4]), (b) involves *only one* confirmational concept, and (c) covers *all* SP cases.
- Finally, let's consider Pearl's *causal* approach to SP [11], which can be fit into our mold (see Extras slide 15 for details).
- Pearl's \mathcal{A}^* (a) has *almost* the same logical form as \mathcal{A} (if you think of Pearl's \mathcal{A}^* as an *enthymeme*), but (b) it involves *causal* and not merely probabilistic/confirmational concepts, and (c) because it (*sensu strictu*) requires an extra premise (*viz.*, P and R are "causally independent," in Pearl's sense), it does *not* apply to *all* cases of SP. [Note: I do not mean to deny that Pearl's story has explanatory value.]

Branden Fitelson Confirmation, Causation, and Simpson's Paradox 10

Example of SP ○ Confirmation-Theoretic "Explanation" of SP ○○○○○○○○ Extras ●○○○○ References

- The following *stochastic truth table* [8] represents all possible probability distributions over the $\{P, Q, R\}$ language, *via* the 7 real variables $a, b, c, d, e, f, g \in [0, 1]$.

P	Q	R	$\Pr(\cdot)$
T	T	T	a
T	T	⊥	b
T	⊥	T	c
T	⊥	⊥	d
⊥	T	T	e
⊥	T	⊥	f
⊥	⊥	T	g
⊥	⊥	⊥	$1 - (a + b + c + d + e + f + g)$

- *In general*, an SP reversal (like our GSA reversal) occurs when the following three inequalities are satisfied.

$$\Pr(Q \mid P \ \& \ R) < \Pr(Q \mid \bar{P} \ \& \ R)$$

$$\Pr(Q \mid P \ \& \ \bar{R}) < \Pr(Q \mid \bar{P} \ \& \ \bar{R})$$

$$\Pr(Q \mid P) > \Pr(Q \mid \bar{P})$$

Branden Fitelson Confirmation, Causation, and Simpson's Paradox 11

Example of SP ○ Confirmation-Theoretic "Explanation" of SP ○○○○○○○○ Extras ○●○○○ References

- Algebraically [8], these three SP inequalities become:

$$\frac{a}{a+c} < \frac{e}{e+g}$$

$$\frac{b}{b+d} < \frac{f}{1 - (a + b + c + d + e + g)}$$

$$\frac{a+b}{a+b+c+d} > \frac{e+f}{1 - (a + b + c + d)}$$
- It is easy to use PrSAT [8] to find *instances* of this pattern (or *any* SP pattern). But, giving a general characterization is quite complex (although, in principle, it is decidable).
- In any case, this does constitute a *general*, algebraic characterization of (dichotomous) SP reversals (which subsumes the case involving statistical frequencies, expressible in terms of 2×2 contingency tables).
- The next two slides contain (algebraic) proofs of the validity of \mathcal{A}_1^* and \mathcal{A}_2^* (using this same algebraic setup).

Branden Fitelson Confirmation, Causation, and Simpson's Paradox 12

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○○○○○	Extras ○○●○○	References
<ul style="list-style-type: none"> ● <i>Probabilistically</i>, \mathcal{A}_1^* is (by definition) as follows. <div style="margin-left: 20px;"> $(1) \Pr(Q P \ \& \ R) > t.$ </div> $(\mathcal{A}_1^*) \frac{(2) \Pr(Q P \ \& \ \bar{R}) > t.}{\therefore (3) \Pr(Q P) > t.}$ ● <i>Algebraically</i> (using the above setup [8]), \mathcal{A}_1^* becomes: <div style="margin-left: 20px;"> $(1) \frac{a}{a+c} > t.$ </div> $(\mathcal{A}_1^*) \frac{(2) \frac{b}{b+d} > t.}{\therefore (3) \frac{a+b}{a+b+c+d} > t.}$ ● Cross-multiplying (1) & (2) yields: <div style="margin-left: 20px;"> $(1) \quad \quad \quad a > (a + c) \cdot t$ $(2) \quad \quad \quad b > (b + d) \cdot t$ </div> ● Adding the lhs & rhs of these and collecting t yields (3). □ <div style="margin-left: 20px;"> $(3) \quad \quad \quad a + b > (a + b + c + d) \cdot t$ </div> 			13

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○○○○○	Extras ○○○●	References
<ul style="list-style-type: none"> ● Here is Pearl's argument form: $\mathcal{A}_{\mathcal{P}}^*$. <div style="margin-left: 20px;"> $(0) \Pr(R do(P)) = \Pr(R do(\bar{P})) = \Pr(R).$ </div> $(\mathcal{A}_{\mathcal{P}}^*) \frac{(1) do(P) \text{ confirms}_i Q, \text{ on the supposition that } R.}{(2) do(P) \text{ confirms}_i Q, \text{ on the supposition that } \bar{R}.}$ $\therefore (3) do(P) \text{ confirms}_i Q, \text{ unconditionally.}$ ● If we think of it as an <i>enthymeme</i> — with premise (0) left <i>unstated</i> — then its form is <i>almost</i>¹⁰ the same as \mathcal{A}. ● Premise (0) asserts that P is <i>casually independent</i> of R. ● In GSA, premise (0) says that <i>intervening on an applicant's gender would not affect the probability that she applies to department E, as opposed to department \bar{E}.</i> ● That seems right. But, there are cases in which (0) <i>fails</i>, but SP occurs (and can still seem, to some extent, "paradoxical"). <p>¹⁰In (1)-(3), $do(P)$ needs to be contrasted with $do(\bar{P})$, not $\bar{do}(P)$, and so the actual form of Pearl's $\mathcal{A}_{\mathcal{P}}^*$ is slightly different than what I've written here.</p>			15

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○○○○○	Extras ○○●○○	References
<ul style="list-style-type: none"> ● <i>Probabilistically</i>, \mathcal{A}_2^* is <i>equivalent to</i>⁹ the following. <div style="margin-left: 20px;"> $(1) \Pr(Q P \ \& \ R) > \Pr(Q).$ </div> $(\mathcal{A}_2^*) \frac{(2) \Pr(Q P \ \& \ \bar{R}) > \Pr(Q).}{\therefore (3) \Pr(Q P) > \Pr(Q).}$ ● <i>Algebraically</i> (using the above setup [8]), \mathcal{A}_1^* becomes: <div style="margin-left: 20px;"> $(1) \frac{a}{a+c} > a + b + e + f.$ </div> $(\mathcal{A}_2^*) \frac{(2) \frac{b}{b+d} > a + b + e + f.}{\therefore (3) \frac{a+b}{a+b+c+d} > a + b + e + f.}$ ● Cross-multiplying (1) & (2) yields: <div style="margin-left: 20px;"> $(1) \quad \quad \quad a > (a + c) \cdot (a + b + e + f)$ $(2) \quad \quad \quad b > (b + d) \cdot (a + b + e + f)$ </div> ● Adding the lhs & rhs of these and collecting t yields (3). □ <div style="margin-left: 20px;"> $(3) \quad \quad \quad a + b > (a + b + c + d) \cdot (a + b + e + f)$ </div> <p>⁹This formulation allows us to prove \mathcal{A}_2^* in <i>the same way</i> we proved \mathcal{A}_1^*.</p>			14

Example of SP ○	Confirmation-Theoretic "Explanation" of SP ○○○○○○○○○	Extras ○○○○	References
<ul style="list-style-type: none"> [1] R. Carnap, <i>Logical Foundations of Probability</i>, 1962. (http://fitelson.org/carnap/logical_foundations_of_probability.pdf) [2] N. Cartwright, "Causal laws and effective strategies," 1979. (http://fitelson.org/Cartwright_CLAES.pdf) [3] V. Crupi, B. Fitelson and K. Tentori, "Probability, confirmation, and the conjunction fallacy," 2008. (http://fitelson.org/pccf.pdf) [4] I. Douven, <i>The Epistemology of Indicative Conditionals</i>, 2016. (http://tiny.cc/mir8cy) [5] B. Fitelson and J. Hawthorne, "The Wason Task(s) and the Paradox of Confirmation," 2010. (http://fitelson.org/wason.pdf) [6] B. Fitelson, "Comments on Khoo & Mandelkern," 2017. (http://fitelson.org/kmc.pdf) [7] _____, "Two New(ish) Triviality Results for the Indicative Conditional," 2016. (http://fitelson.org/triviality_handout.pdf) [8] _____, "A decision procedure for probability calculus with applications," 2008. (http://fitelson.org/pm.pdf) [9] A. Gillies, "Indicative Conditionals," 2014. (http://fitelson.org/gillies_ic.pdf) [10] G. Malinas and J. Bigelow, "Simpson's Paradox," 2016. (http://plato.stanford.edu/archives/sum2016/entries/paradox-simpson/) [11] J. Pearl, "Simpson's Paradox: An Anatomy," 2011. (http://fitelson.org/pearl_sp.pdf) 			16