Types, Tokens and Probabilistic Causation

- Administrative: (i) please post for this week, (ii) volunteer for Thursday, (iii) "minutes" from last 2 discussions TBD, (iv) paper topics.
- Brief Introduction to Probabilistic Causation (more next week)
  - "Causes raise the probabilities of their effects"
  - Correlation $\neq$ Causation [Simpson's Paradox]
  - Homogeneity, causal background contexts, etc.
- Token and Type Probabilistic Causation
  - Sober, Eells, *et al* on tokens and types
  - Good's Example
  - Cartwright's Example
  - Hitchcock's synthesis

Brief Introduction to Probabilistic Causation I

- The basic idea behind probabilistic causation is that "causes raise the probabilities of their effects". Naively, one might then try:

(1) C causes E iff $\Pr(E \mid C) > \Pr(E)$.

- But, (1) won't do, because there are *spurious correlations*. In the early 70's there was a positive correlation between being female (F) and being rejected from Berkeley's graduate school (R). That is, $\Pr(R \mid F) > \Pr(R)$.
- This (initially) raised some suspicions about the possibility of sexual discrimination in the admissions process for Berkeley's grad school.
- If we *partition* the applicants according to the department to which they applied: $\{D_1, \ldots, D_n\}$, then the correlation disappears!
- That is, $\Pr(R \mid F \,\&\, D_i) = \Pr(R \mid D_i)$, for all i. This phenomenon is known as *Simpson's Paradox*. Should we still be suspicious about sexual discrimination? Or, should we view the overall correlation is *spurious*?

Brief Introduction to Probabilistic Causation II

- Simpson's paradox forces us to be much more careful about how we formulate probabilistic accounts of causation. Usually, philosophers move to something like the following more elaborate account:

(2) C causes E iff $\Pr(E \mid C \,\&\, G_i) > \Pr(E \mid G_i)$.

- Here, $\{G_1, \ldots, G_n\}$ form a *partition* (of the relevant probability space), and the each of the G_i is a "homogenous causal background context." Next week, we'll discuss how this partition is typically constructed.
- Some authors (Eells) require that (2) hold for *all* G_i (unanimity). Others (Skyrms, Sober) only require *pareto*-unanimity ($\geq$ for all, $>$ for some). Still others (Dupré) require that (2) hold for a "weighted majority". We will discuss this controversy further next week.
- So far, we have said nothing about what the *relata* of the probabilistic causal relation are. This is where the type/token controversy comes in…

Types and Tokens I

- Eells intends his (2)-theory of probabilistic causation to apply only to type-level causal claims, and not token-level causal claims. Consider:

(3) Smoking causes lung cancer in humans. [type]

(4) David's smoking caused him to develop luncg cancer. [token]

- Sober and Eells both think that the "probability raising" idea is only appropriate for type-level causal claims and not token-level causal claims. Both adopt independent theories of type and token causation.
- They appeal to two kinds of examples to make their case.
  - Good's example and Eells' example
  - Cartwright's example
- Both examples seem to show that we have divergent intuitions about causation and probability at the type *versus* token levels …

Types and Tokens II — Good's Example

Moriarty has perched a boulder on the edge of the cliff, so that when he pushes it carefully it will have a ninety percent chance of killing Holmes. Holmes' intrepid companion Watson arrives atop the cliff just as Moriarty is about to push the boulder. Watson cannot see Holmes from his vantage point, so he is not able to push the boulder in such a way that it will be certain to miss Holmes. Nonetheless, Watson reasons that it is better for him to push the boulder in a random direction than to let Moriarty push it while aiming carefully. Acting quickly on this line of reasoning, Watson rushes forward and pushes the boulder off the cliff in such a way that Holmes chance of dying is reduced to just ten percent. As fate would have it, the boulder crushes Holmes anyway.

- Here, we seem to have the following two conflicting intuitions:

(5) Pre-emptive boulder pushing (P) prevents death by crushing (D).

(6) Watson's pushing the rock (Pw) caused Holmes to die (Dh).

- The (clear) idea behind (5) is that $\Pr(D \mid P) < \Pr(D \mid \neg P)$. But, why do we think that (6) is true? Influence? Counterfactual dependence?

Types and Tokens III — Cartwright's Example

Nancy discovers poison oak in her garden, and she decides to spray it with a defoliant. The defoliant claims to be ninety percent effective, meaning that if properly used, the poison oak has a ninety percent chance of dying within a specified period of time – say one month. Let us assume that this claim is accurate. Let us assume, moreover, that if left untreated, the poison oak would have only a ten percent chance of dying within one month. Thus, spraying with this defoliant decreases the chances of survival for plants of this kind from ninety percent to ten percent. As it happens, however, the plant survives.

- Here, we seem to have:

(7) Defolaint spraying (S) prevents poison oak survival (O).

But, we would (presumably) *not* want to say:

(8) Nancy's spraying her poison oak (Sn) caused it to survive (Op).

- Cartwright's example has the same probabilistic structure as Good's example. Yet, they seem to have different token-level analyses.

Types and Tokens IV — Hitchcock's Synthesis I

- Hitchcock suggests the following "neo-Humean" account of token probabilistic causation in terms of type-level probabilistic causation:

($\mathcal{H}$) x's being C caused x's being E just in case (i) x was C, (ii) x became E, and (iii) C causes E (in the type-level p-causal sense).

- Examples such as Good's and Cartwright's are *prima facie* counterexamples to $\mathcal{H}$, if (iii) is unpacked in the standard way.
- Hitchcock suggests that we understand causal claims as being of the form C *in contrast to* C' causes E. In Good's example, if N is that nobody pushed the boulder, and Pm is that Moriarty pushed the boulder, then:

(9) Pw, in contrast to N, caused Holmes' death.

(10) Pw, in contrast to Pm, prevented Holmes' death.

- This allows us to explain our conflicting intuitions about Good's example in a way that does *not* depend on the type/token distinction.

Types and Tokens V — Hitchcock's Synthesis II

- In Good's example, there is a salient alternative (nobody pushing the rock, N) in which Holmes has virtually no chance of dying.
- In Cartwright's example, on the other hand, there is no salient alternative in which the poinson oak has virtually no chance for survival.
- This is why (according to Hitchcock) Cartwright's example is not analogous to Good's example. But, change Cartwright's example a bit.
- Say Nancy *chooses* the 90% effective defoliant over a 99% effective defoliant (say, because it was cheaper). Then there is a salient alternative (99% defoliant, $S'n$) under which the plant will almost surely die. Then,

(11) Sn, in contrast to $S'n$, caused the plant to survive.

(12) Sn, in contrast to N (no defoliant sprayed), prevented the plant's survival.

- This allows us to restore the analogy, and provide a unified analysis of the two examples, without appealing to the type/token distinction.

Types and Tokens VI — Hitchcock's Synthesis III

- It seems to me that in Good's example, it is part of the background (causal field $\mathfrak{F}$) that if Watson had not pushed the rock, Moriarty would have. But, if that's right, then how is Hitchcock's maneuver possible?
- Hitchcock claims that, if H = Holmes dies, W = Watson pushes the rock, N = nobody pushes the rock, M = Moriarty pushes the rock, then $\Pr(H \mid W) < \Pr(H \mid M)$, and $\Pr(H \mid W) > \Pr(H \mid N)$.
- But, as I understood the example, $\mathfrak{F}$ implies $\neg W \,\Box\!\!\rightarrow M$. If that's right, then doesn't $\mathfrak{F}$ imply $\neg N$? If so, then $\Pr(H \mid N \,\&\, \mathfrak{F})$ is *undefined*, so $\Pr(H \mid W \,\&\, \mathfrak{F}) \not> \Pr(H \mid N \,\&\, \mathfrak{F})$. If $\mathfrak{F} \not\models \neg N$, why not?
- In general, I like Hitchcock's move here. It forces the *p*-causal relation to be explicitly *contrastive*. And, *p*-causation is already explicitly *contextual* (since Pr functions must presuppose a *reference class*).
- That brings us to the problem of the reference class, and the problem of single-case or token-event probabilities (two key problems in probability).

Types and Tokens VI — Hitchcock's Synthesis IV

- Hitchcock gives an ingenious formal apparatus for simultaneously characteriaing token and type level probabilistic causal claims.
- Let $\langle \Omega, \mathcal{F}, \Pr \rangle$ be a probability space, $X_1, X_2, \ldots$ be a sequence of random variables on the space, and $E_1, E_2, \ldots$ be a sequence of events in $\mathcal{F}$.
- For instance, let the indexing correspond to individuals: David = 1, Mary = 2, etc. And, let E_i correspond to i's developing lung cancer, and $X_i = n$ represent i's smoking n packs of cigarettes per day.
- Now (assuming other causal factors are held fixed), claims about the causal relevance of i's smoking for i's developing lung cancer describe the conditional probability function $f_i(x) = \Pr(E_i \mid X_i = x)$.
- David's smoking 2 packs per day (as opposed to not smoking at all) caused him to develop lung cancer: $\Pr(E_1 \mid X_1 = 2) > \Pr(E_1 \mid X_1 = 0)$.
- So, that gives us a way to handle *token* (contrastive) causal claims.

Types and Tokens VII — Hitchcock's Synthesis V

- Hitchcock shows that it is possible to construct a "general" probability space $\langle \Omega_i, \mathcal{F}_i, \Pr_i \rangle$ such that E is in and event in $\mathcal{F}_i$, and X is a random variable over Ω_i such that $f_i(x) = \Pr_i(E \mid X = x)$.
- And, if the conditional probability functions $f_i(x)$ are identical, then the same space can be constrructed for each one. That is, we can construct a space $\langle \Omega', \mathcal{F}', \Pr' \rangle$ such that $\Pr'(E \mid X = x) = \Pr(E_i \mid X_i = x)$, for all i.
- Analogy: Let Ω consist of sequences of H's and T's (coin tosses). And, let $X_1, X_2, \ldots$ be a sequence of random variables such that for each $\omega \in \Omega$, $X_i(\omega)$ is the ith term in sequence ω. Thus, the event $X_8 = T$ corresponds to to the set of all sequences ω that have a T in the eigth position.
- Now, the *distribution function* $\mu_i(T) = \Pr(X_i = T)$ will itself be a perfectly good probability measure. And, if the X_i are assumed to be identically distributed (as they are with coins), then we will have a single distribution function μ, which gives "type-level" probabilities.

Types and Tokens VIII — Hitchcock's Synthesis VI

- Returning to Hitchcock's analysis, he assumes that the $f_i(x)$ are identical, which allows him to construct a single "type-level" condiotional probability function $\Pr'(E \mid X = x)$. He calls $\langle \Omega, \mathcal{F}, \Pr \rangle$ the "big" probabiloity space, and $\langle \Omega', \mathcal{F}', \Pr' \rangle$ the "little" probability space.
- This involves a *uniformity of nature* assumption: if two indioviduals i and j are alike with respect to all of the factgors that are causally relevant to E, then i and j will not have a different probability of E, simply in virtue of being i and j (analogy to coins, probability of heads).
- Generalization: For each i, there will be a partition of possible background contexts into cells $\{G_i^1, G_i^2, \ldots\}$. It seems plausible that the functions $f_{ij}(x) = \Pr(E_i \mid X_i \,\&\, G_i^j)$ will be independent of i.
- To simplify notation, let $\mathbf{Y}_i$ be a random vector such that $\mathbf{Y}_i = (x, j)$ is identical to $X_i = x \,\&\, G_i^j$. Then, Hitchcock shows that we can construct a $\Pr'$ such that $\Pr'(E \mid \mathbf{Y} = \mathbf{y}) = \Pr(E_i \mid \mathbf{Y}_i = \mathbf{y})$, if uniformity holds.

Reference Classes and Single/Token Cases

- What is the probability that John Doe will get cancer in the next year? Well, that depends on what *reference class* you have in mind. If you choose a reference class too narrowly (things identical to John Doe), then it seems that the probability (relative to this reference class) is either 1 or 0 (depending on whether he does in fact contract cancer).
- If we choose too broadly, then we're leaving out relevant factors. However, if we take probability to be a propensity of John Doe himself (given his life-history up to this point, and everything about him), then we have no trouble thinking about the probability of this token event. This is just a disposition that John possesses (in his current context).
- This is importantly different than the question "What is the probability that a 47 year old male with a 20-pack-year smoking habit will get cancer in the next year?" This is a type-level question about relations between properties, within a population of a certain kind.