**Causal networks.** Preemption seems to show that there can be counterfactual dependence without causation. But if one looks at examples of preemption, it often seems that \( e \) does in a way depend on \( c \). It depends on \( c \) “in the obtaining circumstances” or “holding certain things fixed.”

*Early preemption.* Suzy and Billy both throw their rocks at a bottle. Suzy’s rock deflects Billy’s en route to the bottle, and Billy’s rock falls to the ground. True, it’s not the case that \( \sim O_c \rightarrow \sim O_e \). But it is the case that there are facts \( G \) such that \( \sim O_c \& G \rightarrow \sim O_e \). Name some.

*Late preemption.* Same setup, except that Billy’s rock is not deflected; it just arrives too late, after the bottle has been broken by Suzy’s rock. True, it’s not the case that \( \sim O_c \rightarrow \sim O_e \). See if you can think of a \( G \) such that \( \sim O_c \& G \rightarrow \sim O_e \).

*Trumping preemption.* The troops obey Colonel and they obey Captain, but if both make an order simultaneously, it’s the higher ranking officer (Captain) they’ve been trained to follow. Colonel and Captain both shout “March!” The troops march because of Captain’s order. But if Captain had kept quiet, they still would have marched. What can we hold fixed to restore the dependence?/

*Magical preemption.* The laws of magic say the first spell of the day determines what happens at midnight. Merlin says in the morning, “Prince into frog!” Morgana says the same thing in the evening. It’s Merlin’s spell that causes the transformation, but there’s no counterfactual dependence. What to hold fixed?

*Hastening preemption.* The forest would have burned in June anyway due to a big lightning storm, but the drought dries it out so that a small lightning storm in May is enough to set it off. One wants to say the drought is a cause of the fire, despite that it would have occurred anyway. What to hold fixed?

One implementation of the holding fixed strategy has arisen in engineering, in the work of Judea Pearl (*Causality*, 2000) and others. It’s called the method of causal networks. Hitchcock gives a simplified version of it. The basic idea is that \( c \) causes \( e \) iff there’s an “active” path or route through the relevant network, leading from \( c \) to \( e \).

First we explain networks. One starts with a bunch of variables \( A, B, C \), etc. To keep things simple they can only assume two values: 1 if the corresponding event \( a, b, c, \ldots \) occurs, 0 if it does not occur. So, \( C \) might be Suzy’s throwing, \( D \) might be Billy’s, and \( E \) might be the bottle shattering. One states the counterfactual relations between the
values of these variables using “structural equations.” Assume that Billy throws only if Suzy doesn’t. Then the equations are

\[ D = \neg C \]
\[ E = C \lor D \]

Read these as convenient shorthand for counterfactuals. The first says that \( C=0 \rightarrow D=1, \quad C=1 \rightarrow D=0; \) the second says that \((C=1 \lor D=1) \rightarrow D=1, \quad (C=0 \land D=0) \rightarrow E=0.\) The equations tell us how to interpret this graph; they tell us how arrow-tip variables have their values determined by arrow-base variables.

\[ \begin{array}{c}
C \\
\downarrow \\
D \\
\uparrow \\
E
\end{array} \]

And now here is what is meant by an active path. To determine if there is an active path from C to E, one first sets off-path variables to their actual values. That means one changes the equation for D to this: \( D=0 \) (since we’re assuming Billy didn’t throw). One now toggles C up and down from 0 to 1 and back and looks at how this affects E. Clearly, if we hold D fixed at 0, then E is just going to take the same value as C. The path is active insofar as changing C’s value (with off-path variables frozen at their actual values) changes E’s value. The claim (slightly simplified) is that \( c \) causes \( e \) iff there is an active path from C to E, so that toggling C on and off is a way of toggling E on and off.

One nice feature of this account is that it fits with our intuitions about transitivity. Suppose \( c \) is a boulder being dislodged, \( d \) is the hiker seeing it dislodged and ducking, and \( e \) is the hiker’s surviving. The equations are \( D=C, \quad E=\neg C \lor D; \) the diagram can be the same as above. Freezing D to its actual value of 1 (the hiker ducked) the route from C to E is inactive, even though the routes from C to D and from D to E are both active. The route via D is inactive too; the hiker survives whether the boulder is dislodged or not.

Problem, though: suppose we add a new variable \( D^* \) on the upper path, equal to 1 if there’s a falling boulder right by the hiker’s head, otherwise 0. The equations are \( D^*=C, \quad D=C, \quad E=D \lor \neg D^*.\) Now (exercise!) the path from C to E through D becomes active. Uh oh! Holding fixed that there was a boulder about to fall on her head, it’s a good thing a boulder was dislodged a moment earlier, leading the hiker to duck! The bomb example is even clearer in this respect. Holding fixed that the chair was going to explode, whether the victim survives depends on whether a bomb is put under her chair to tip her off to the impending explosion! Next time time talk about responses to these counterexamples.