# Lecture 2
# Agency and Effective Strategies

## 1. Introduction & Gasking

Last week we looked at some canonical positions in the metaphysics of causation. We also looked at several different kinds of recalcitrant case that challenged those theories (e.g. preemption cases, transitivity, causation by omission).

One motivation for considering an agency account comes from its purported ability to avoid those counterexamples. Another comes from the ways in which we investigate and reason about causes.

One of the earliest examples of an agency-based theory comes from Gasking (1955) (Collingwood (1940) also wrote on similar ideas). Gasking writes "a statement about the cause of something is very closely connected with a recipe for producing it or for preventing it" (483).

Gasking considers the following hypothetical case:

> *Our early ancestors many millennia ago discovered that you could make a large range of substances (wood, water, leaves, etc.) glow first blue, then purple, then red by a process of alternately covering them so as to exclude light, then rapidly letting light fall on them, then quickly covering them again, and so on. […] Then, about 1000 B.C. men got hold of samples of fairly pure iron, for the first time. They tried the covering- uncovering technique on it to see if it […] would glow if manipulated in this way. They found that it would, but that, unlike other substances, iron began to get hot when it started glowing, got hotter still at the purple stage, and when glowing red was very hot indeed. Precise measurements in modern times showed that on reaching the red stage the temperature of iron was 1,000C. In other respects this imaginary world is just like our world, except that when you put a poker or other non-combustible object in a fire it does not begin to glow, however hot it gets.* (481)

He also notes some of the practical uses of our current causal discourse—for instance, we often use causal claims as a way of telling someone how they may make something happen. E.g. "If you want the lights on, flick the switch furthest to the left."

He eventually settles on the following account: "one says 'A causes B' in cases where one could produce an event or state of the A sort as a means to producing one of the B sort" (485).

## 2. Cartwright on Effective Strategies

While Cartwright's name is often spoken in the same breath as other manipulationists, it's important to note that she's not herself a manipulationist. That said, insights from her 1979 paper have greatly influenced the manipulationists that came after.

Cartwright begins with that classic question: what is the difference between *causal* regularities and mere regularity? She argues that the difference lies in which of the regularities are also indicative of some effective strategy. So, purchasing the insurance she mentions is *not* an effective strategy for prolonging one's life, but striking matches is an effective strategy for having lit matches. Indeed, Cartwright argues that, without causation, we *couldn't* demarcate between regularities that are also effective strategies and regularities that are not.

She then gives the following probabilistic characterisation of causation:

> *C causes E* if and only if *C* increases the probability of *E* in every situation which is otherwise causally homogeneous with respect to *E*.
>
> Carnap's [3] notion of a state description can be used to pick out the causally homogeneous situations. A complete set of causal factors for *E* is the set of all $C_i$ such that either $C_i \rightarrow +E$ or $C_i \rightarrow -E$. (For short $C_i \rightarrow \pm E$.) Every possible arrangement of the factors from a set which is complete except for *C* picks out a population homogeneous in all causal factors but *C*. Each such arrangement is given by one of the $2^n$ state descriptions $K_j = \bigwedge \pm C_i$ over the set $\{C_i\}$ (*i* ranging from 1 to n) consisting of all alternative causal factors. These are the only situations in which probabilities tell anything about causal laws. I will refer to them as *test* situations for the law $C \rightarrow E$.
>
> Using this notation the connection between laws of association and causal laws is this:
>
> *CC*: $C \rightarrow E$ iff $P(E/C.K_j) > P(E/K_j)$ for all state descriptions $K_j$ over the set $\{C_i\}$, where $\{C_i\}$ satisfies
> (i) $C_i \in \{C_i\} \Rightarrow C_i \rightarrow \pm E$
> (ii) $C \notin \{C_i\}$
> (iii) $\forall D\ (D \rightarrow \pm E \Rightarrow D = C \text{ or } D \in \{C_i\})$
> (iv) $C_i \in \{C_i\} \Rightarrow \neg(C \rightarrow C_i)$.

The conditions (i)-(iv) can be expressed in natural language thus:

(i) All the members of $\{C_i\}$ causally influence E
(ii) C is not one of the causal factors in $\{C_i\}$
(iii) Anything (other than C) that causally influences E is in $\{C_i\}$
(iv) None of the causal factors in $\{C_i\}$ are effects of C

These are the rules that govern the set of **causal factors** $\{C_i\}$. From these causal factors, we generate what Cartwright calls '**state descriptions**'. To generate a single state description, take each member of $\{C_i\}$ and set it to a particular value (e.g. occurs, does not occur). Of course, the members of $\{C_i\}$ could take different combinations of values, so for any given set $\{C_i\}$, there will be multiple possible state descriptions. Think of these like different kinds of background conditions.

Now, to restate Cartwright's theory, *C causes E* (type-level claim) if and only if, in every single one of the state descriptions we could define over the relevant set of causal factors, *C* raises the probability of *E*.

**Simpson's Paradox**

We know that smoking causes heart disease. We also know that exercise prevents heart disease (i.e. causes one not get heart disease). Suppose, for the sake of the example, that most smokers exercise daily. Suppose also that the causal influence of exercise is stronger than the causal influence of smoking. In this case, if we only looked at probability-raising (without taking care to distinguish background conditions), it would be *false* that smoking raises the probability of heart disease!

*Question*: What do I consider when I am trying to decide what to do? Probability-raising relationships, or causal relationships?

According to Cartwright, you must consider causal relationships. And these cannot be reduced to probability-raising relationships. Suppose for *reductio* that an action is an effective strategy iff it raises the probability of the desired outcome. It follows that, in the example above, smoking is an effective strategy for preventing heart disease. But, in the case as described, smoking is not an effective strategy for preventing heart disease. Therefore, not all actions that raise the probability of the desired outcome are effective strategies.

Cartwright's own account avoids problems arising from Simpson's Paradox (as well as those arising from the other recalcitrant cases we looked at last week). But it does so by including the concept of causation in the characterisation of causation. In other words, her account is **nonreductive**.

## 3. Price on Agency and Decision-Making

Price (1991) agrees with Cartwright that causation and effective strategies are importantly related. However, he disagrees with her insofar as he thinks that we *can* distinguish effective strategies from ineffective ones without appeal to an irreducible concept of causation.

### Coco and the Mars Bar

In the case Price describes, the background facts are these: eating chocolate does not cause migraines; a pre-migrainous state (PMS) causes migraines; PMS causes chocolate consumption. (Further, let's assume that PMS raises the probability of its effects.) Against this backdrop, Coco—who loves chocolate—is deciding whether or not to eat a Mars Bar. A problem seems to arise: if Coco reasons strictly probabilistically, they should choose not to eat the chocolate, because eating the chocolate raises the probability of having a migraine. But, this seems irrational, since eating chocolate doesn't cause migraines.

### Agent Probabilities

In his answer to the Mars Bar problem, Price argues that Coco's deliberations about which action to perform (eat or not-eat) themselves provide the solution. He explains that eating chocolate only raises the probability of getting a migraine when eating chocolate was caused by PMS (and conversely, not eating chocolate only raises the probability of not getting a migraine if the lack of PMS caused the failure to eat chocolate). But, in deliberating about what to do, Coco is deciding whether or not to eat chocolate based on their other beliefs (namely, based on their belief that eating chocolate raises the probability of getting a migraine). If Coco decides not to eat the chocolate, then the failure to eat chocolate will have been caused by their belief, and not caused by PMS. In this situation, the correlation between chocolate-eating and migraines disappears.

From this, Price concludes that the conditional probabilities relevant to effective strategies (and, eventually, causation) are **agent probabilities**: conditional probabilities, where the condition is conceived of as brought about by a free agent. So, in the Mars Bar case, the conditional probability of getting a migraine given that you eat chocolate is greater than if you don't eat chocolate; however, the conditional probability of getting a migraine given that a free agent decides to eat chocolate is no greater than if the agent decided otherwise.

## 4. Price and Menzies on Secondary Qualities

In this paper, Price and Menzies defend the agency theory of causation from a series of objections by arguing that causation is something like a secondary quality. They do this by drawing an analogy between the agency theory of causation and the dispositional account of colour.

Menzies and Price state their theory thus: *A is a cause of B just in case, from the perspective of a rational agent, A would be an effective means of bringing about B*.

### Obj. 1 Epistemology confused with metaphysics

The agency theory confuses the way we *discover* causes with what it *is* to be a causal relationship. Just because I *learn* that striking a match causes it to light by striking matches, it doesn't follow that the causal relationship itself is constituted by my being able to make a match light by striking it.

Against this objection Menzies and Price argue that they take causes to be secondary qualities in much the same way colours are (on a dispositional theory), and as such maintain the property of being causal that some relations share is, in fact, an *extrinsic property*. Just as what it is to be red is to be some relation between an object, a normal observer, and normal conditions, what it is for a relation to be causal is to be some relation between an agent and two distinct events.

### Obj. 2 Vicious Circularity

The agency theory invokes a *bringing about* relation between the agent and the putative cause A. But bringing about is a *causal* relation. Thus the account analyses causation in terms of causation, and so is viciously circular.

Menzies and Price argue that, in fact, the theory is not circular, despite appearances; the theory does not need to refer to bringing about per se but to the *experience* of bringing about.

### Obj. 3 Unmanipulable Events

There are many cases where we believe A is a cause of B, but we are unable to manipulate A (e.g. the movement of tectonic plates causes earthquakes). The agency theory looks compelled to say that, in cases like this, A is not a cause of B.

In response, Menzies and Price argue that we can make inferences from events that we can manipulate to those that we cannot by reasoning by analogy (provided the two sets of events are similar with respect to certain intrinsic features). So, for instance, I might build a scale model of tectonic plates and buildings, then manipulate the model by moving the model plates against one another. When the model buildings topple, and the model ground cracks, this gives me reason to think that the movement of the Earth's tectonic plates was a cause of the earthquake.

**Obj. 4 Unacceptable Anthropocentricity**

Version 1 – Intuitively, there would have been causal relations even if agents did not exist; but the agency theory seems to render this false, since the account depends on there being agents to bring about events.

In response, Menzies and Price note that even at an agent-less world, it would still be the case that *if* an agent brought about A, B would occur.

Version 2 – On the agency theory, if our manipulative abilities had been different, the causal relations would have been different. For instance, suppose at another world, our physical abilities are so restricted that we could only move our eyes in their sockets. It would seem to follow that dropping glass does not cause glass to shatter since we could not bring about the dropping.

Menzies and Price respond by accepting that this is a consequence of their account but deny that it is problematic. It is no more problematic, they argue, than the fact that if it were the case that our visual sense organs were different, different objects would have been red, or if it were the case that our taste receptors were different, different foods would have been sweet. This is simply what it is to be a secondary quality.