9.1. A Response to Prelec

Luc Bovens

Outcome and diagnostic utility

At the heart of Drazen Prelec's chapter is the distinction between outcome utility and diagnostic utility. This distinction becomes interesting when a strict regard for outcome utility prompts us to do one thing, whereas taking into account diagnostic utility prompts us to do something else. Let us look at Prelec's paradigm cases.

(i) A Calvinist is considering whether to engage in a single sinful (and pleasurable) action.
(ii) A person who has quit smoking is considering whether to smoke just one single cigarette.
(iii) Rather than investing in a scheme that benefits orphans, a nation is considering whether to make a different investment that has greater public benefits (say roadworks).

The outcome utility of the single sinful action, smoking the single cigarette and making a different investment is positive, but the diagnostic utility is negative. The diagnostic utility of the single sinful action is the utility that is associated with not being among God's chosen; the diagnostic utility of smoking the single cigarette is the utility associated with being a weak-willed person who won't be able to stick to her resolution of quitting smoking; and the diagnostic utility of choosing the different investment is the utility associated with being the kind of society that is heartless and does not care about the needy. A neo-Calvinist decision-maker will refrain from the actions in (i), (ii) and (iii) on the grounds of their negative diagnostic utility, although their outcome utility is positive.

Observing and intervening in causal networks

There is a particular distinction in the literature on causal networks, namely the distinction between *observing* and *intervening*, that maps onto Prelec's distinction between diagnostic and outcome utility. I will explore the connection between both frameworks.

Here is a slightly simplified version of Pearl's paradigm example (Pearl, 2000, p. 15). Let there be the binary variables season (*winter* or *summer*), rain (*yes* or *no*), sprinkler (off or on) and wet (*yes* or *no*). The chance of rain (in California) is greater in winter than in summer. People tend to turn their sprinklers on more often in summer than in winter. Rain and sprinklers tend to make the pavement wet, though rain's effect is substantial while the sprinkler's effect is marginal. Figure 9.1.1 represents this causal structure and if we insert low values for δ and ε, then the probabilities in the figure match our description of weather patterns and sprinkler settings in Californian seasons and the effect of rain and sprinklers on pavements.

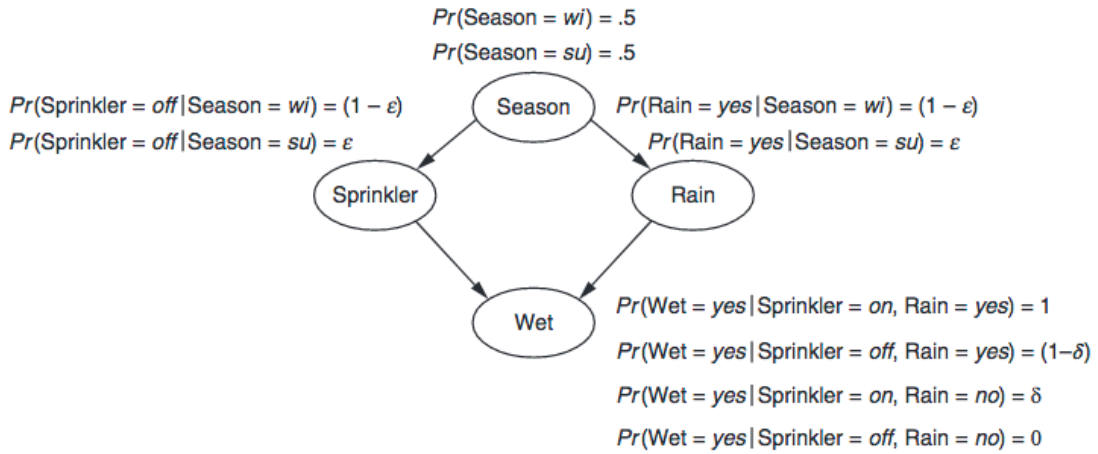

Figure 9.1.1 Pearl's sprinkler network

Let us slightly change the example to bring in utilities. Suppose that the water does not drip onto the pavement but onto a bush and that the utility of a wet bush (i.e. a bush that is sufficiently wet to flower) is 1 and of a dry bush (a bush that is insufficiently wet to flower) is 0. Hence the expected utility equals the probability of Wet = *yes*.

What is the change in expected utility of *observing* that someone has (rather than has not) turned on the sprinkler? Let us set ε as approaching 0 and δ at 1/6 in the graph. Winter makes it very likely that there is rain and that the sprinkler is off, whereas summer makes it very likely that there is no rain and that the sprinkler is on. Rain and sprinkler action provide certainty that the bush will flower, rain by itself a 5/6 chance, sprinkler action by itself 1/6 chance and no rain or sprinkler action provide certainty that the bush won't flower. Now suppose that we observe that someone has turned on the sprinkler. Then we can infer that it is very likely to be summer. Hence it is very likely that there is no rain and so the expected utility is (roughly) (1 – δ) = 1/6. Suppose that we observe that nobody has turned on the sprinkler. Then we can infer that it is very likely to be winter. Hence there is likely to be rain and the expected utility is (roughly) (1 – δ) = 5/6. So *observing* that the sprinkler is on rather than off entails a drop in expected utility of (1 – δ) – δ = (5/6 – 1/6) = 2/3.

What is the change in expected utility of *intervening* by turning on the sprinkler (rather than not turning it on)? If it's summer, then the expected utility of turning on the sprinkler is 1/6 and of not turning on the sprinkler is 0. If it's winter, then the expected utility of turning on the sprinkler is 1 and of not turning on the sprinkler is 5/6. So either way, *intervening* by turning on the sprinkler (rather than leaving it turned off) entails a rise of expected utility of δ = 1/6.

In Prelec's terms, the diagnostic utility of the sprinkler being on rather than off is negative: *observing* that the sprinkler is on rather than off is bad news since it entails a drop in expected

utility. However, the outcome utility of the sprinkler being on rather than off is positive: *intervening* to turn the sprinkler on is a good thing to do since it entails a rise in expected utility.

I will show now how Prelec's cases of neo-Calvinism have the same structure as Pearl's paradigm example.

Let us start with the case of the Calvinist. Committing a single sinful action is like the sprinkler being turned on. Being destined to heaven is like rain. Being among God's chosen is like wintertime in California. We can simply substitute these new variables into Figure 9.1.1. *Mutatis mutandis* we can now rewrite the two paragraphs above in which I introduced the distinction between *observing* and *intervening*.

What is the change in expected utility of *observing* myself committing a single sinful action rather than refraining? Suppose that I observe myself committing such an action. Then I can infer that I am very likely not to be among God's chosen. Hence it is very likely that I am not destined for heaven and so my expected utility is (roughly) $\delta = 1/6$. Suppose that I observe myself refraining from this single sinful action. Then I can infer that I am very likely to be among God's chosen. Hence it is likely that I am destined for heaven and so my expected utility is (roughly) $(1 - \delta) = 5/6$. So *observing* myself committing a single sinful action rather than refraining entails a drop in expected utility of $(1 - \delta) - \delta = (5/6 - 1/6) = 2/3$.

What is the change in expected utility of *intervening* by committing a single sinful action rather than refraining? If I am not among God's chosen, then the expected utility of committing this single sinful action is 1/6 and of refraining is 0. If I am among God's chosen, then the expected utility of committing this single sinful action is 1 and of refraining is 5/6. So either way, *intervening* by committing the single sinful action entails a rise to expected utility of $\delta = 1/6$.

Hence the diagnostic utility of committing a single sinful action is negative but the outcome utility is positive. Prelec's two other cases can be dealt with in the same fashion.

Here is the smoking case: smoking a single cigarette is like the sprinkler being turned on; refraining from future smoking is like rain; being wilful is like wintertime in California. *Observing* myself smoking a single cigarette entails a drop in expected utility since I infer that it is very likely that I am not wilful and won't be able to refrain from smoking in the future. *Intervening* by smoking a single cigarette entails a rise in expected utility. The diagnostic utility of smoking a single cigarette is negative but the outcome utility is positive.

Here is the orphan scheme case: refraining from investing in an inefficient orphan scheme on the grounds that another investment has greater public utility is like the sprinkler being turned on; investing in caring schemes at large is like rain; being a caring society is like wintertime in California. *Observing* my society refrain from investing in the orphan scheme entails a drop in expected utility since I infer that it is very likely that we are not a caring society and that we won't invest in caring schemes at large. *Intervening* in my society by refraining from investing in the orphan scheme and investing in a scheme with greater public benefit entails a rise in

expected utility. The diagnostic utility of refraining from investing in the orphan scheme is negative but the outcome utility is positive.

Prelec takes the agent's total utility to be her outcome utility complemented by her diagnostic utility. Then, following our calculation above, the total utility of turning the sprinkler on, committing the single sinful action ..., equals a δ-gain in outcome utility and a drop of ((1 –δ) – δ) in diagnostic utility: δ – ((1 – δ) – δ). And vice versa, the total utility of turning the sprinkler off, refraining from a single sinful action ..., equals a δ-drop in outcome utility and a gain of ((1 – δ) – δ) in diagnostic utility: – δ + ((1 – δ) – δ). This is somewhat different from Prelec's de finition of total utility, but it is the closest that I can come to it.[1]

Interpretation

I take it that Prelec aims to construct a model of an agent who is partly guided by outcome utility and partly by diagnostic utility. Whether this agent is rational or not is a non-issue. This model has descriptive value–it captures the decisions of actual human agents facing problems that have this kind of structure. This is a valid pursuit. It is comparable say to the model in prospect theory showing that agents tend to make decisions by underweighting high probabilities and overweighting low probabilities (Kahneman and Tversky, 1979). These are not rational choice explanations. We do not attempt to explain agency by showing that what the agent does is a rational action.

But one might ask another question that does belong to rational choice theory, namely are there any situations under which a rational agent would be advised to leave the sprinkler off, to refrain from committing the single sinful action, etc. In our simple model the answer is a resolute no–i.e. turning the sprinkler on, committing the single sinful action..., is the only rational option. The only thing that matters when deciding whether or not to turn the sprinkler on, to commit the single sinful action ..., is that one raises one's outcome utility by δ = 1/6 by doing so, no matter what season it is, no matter whether one is among God's chosen.[2]

[1] Alternatively, one could calculate diagnostic utility as the change in expected utility from *observing* a third party leaving the sprinkler off rather than being ignorant about whether she has turned the sprinkler on or left it off and representing this ignorance as equiprobability. Then the diagnostic utility of her leaving the sprinkler off raises my expected utility by (1 – δ) – (.5(1 –δ) + .5δ) = 5/6 – 1/2 = 1/3. Again, this does not quite coincide with Prelec's definition.

[2] This is the advice of the causal decision theorist in the Newcomb problem in Nozick (1969). The Newcomb problem has the same structure. Substitute 'me having the character of a two-boxer' for 'summer', 'me taking two boxes' for 'the sprinkler being on' and 'the predictor putting money in the opaque box' for 'rain'. As a causal decision theorist, I take two-boxing to be the only rational solution.

But now, in real life, it may indeed be a good idea to refrain from turning the sprinkler on or to abstain from the single sinful action. So why is this? Well, in real life, the causal structure is often more complex and my singular agency does have causal consequences that go beyond immediate consequences.

In the smoking case, it is reasonable to assume that there is no fixed character. Today's character may be characterized by great resolve, but after smoking one cigarette, my character will lose its resolve. With less resolve comes a reduced chance that I will be able to continue abstaining from smoking. And this is why I should abstain from smoking a single cigarette.

In the Calvinist case, it is reasonable to assume that Calvinists tend to draw inferences from their singular actions about their chances of salvation and these inferences very much influence their peace of mind. A rational agent may need to work with such projected beliefs and if we include the agent's peace of mind in the description of the states of the world over which preferences are defined, then a rational agent may indeed need to abstain from the single sinful action.

But if we want to capture such features in our model of rational agency, then we need to construct a more complex causal network that includes downstream nodes from committing a single sinful action. For example, in our smoking case, this would be the future state of my character and its effect on long-term smoking. In our Calvinist case, it would be our future cognitive state and the resultant level of peace of mind.

The orphan scheme can be thought of in a similar vein. A rational society might want to invest in the inefficient orphan scheme. The reason for doing so is that turning one's back on such a scheme may make society redefine what kind of society it is. It may come to think of itself as a heartless society. And it will then lose its resolve to invest in caring schemes at large. Or its citizens may derive much benefit from the self-congratulatory state of mind that ensues from investing in highly visible caring schemes (even if they are inefficient). If we import this more complex causal structure into our cases, then it will become rational to invest in the otherwise inefficient orphan scheme.

References

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47(2): 263–92.

Nozick, R. (1969). Newcomb's Problem and Two Principles of Choice. In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel, pp. 114–46.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press