(version of 6/02/01) This is an expanded version of a paper forthcoming in *Synthese*.

Causal Probability

John L. Pollock
Department of Philosophy
University of Arizona
Tucson, Arizona 85721
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

Abstract

Examples growing out of the Newcomb problem have convinced many people that decision theory should proceed in terms of some kind of causal probability. I endorse this view and define and investigate a variety of causal probability. My definition is related to Skyrms' definition, but proceeds in terms of objective probabilities rather than subjective probabilities and avoids taking causal dependence as a primitive concept.

1. Causal Decision Theory

Decision theory is a theory of rational choice. It is a theory of how, rationally, an agent should go about deciding what actions to perform at any given time. The basic ideas of classical decision theory can be stated simply. We assume that our task is to choose an action from a set A of *alternative actions*. The actions are to be evaluated in terms of their outcomes. We assume that the *possible outcomes* of performing these actions are partitioned into a set O of pairwise exclusive and jointly exhaustive outcomes. We further assume that we know the probability $\mathsf{PROB}(O/A)$ of each outcome conditional on the performance of each action. Finally, we assume a *utility-measure* $\mathbf{U}(O)$ assigning a numerical utility value to each possible outcome. The *expected-value* of an action is defined to be a weighted average of the values of the outcomes, discounting each by the probability of that being true if the action is performed:

$$\mathbf{EV}(A) = \sum_{O\in \mathrm{O}} \mathbf{U}(O)\cdot\mathsf{PROB}(O/A).$$

The crux of classical decision theory is that actions are to be compared in terms of their expected-values, and rationality dictates choosing an action that is *optimal*, i.e., such that no alternative has a higher expected-value.

This work was supported by NSF grant no. IRI-IIS-0080888.

Nozick's (1969) presentation of the Newcomb problem led to a general recognition that classical decision theory is flawed, making incorrect prescriptions in some cases. The Newcomb problem itself commands conflicting intuitions, but there are other examples that are clearer. One of the more compelling examples is due to Stalnaker (1978). Suppose you are deciding whether to smoke. Suppose you know that smoking is somewhat pleasurable, and harmless. However, there is also a "smoking gene" present in many people, and that gene both (1) causes them to desire to smoke and (2) predisposes them to get cancer (but not by smoking). Smoking is evidence that one has the smoking gene, and so it raises the probability that one will get cancer. Getting cancer more than outweighs the pleasure one will get from smoking, so classical decision theory recommends against smoking. But this seems clearly wrong. Smoking does not *cause* cancer. It is just evidence that one already has the smoking gene and hence may get cancer from that. If you have the smoking gene, you will still have it even if you refrain from smoking, so the latter will not prevent your getting cancer.

As a number of authors (Gibbard and Harper 1978; Sobel 1978; Skyrms 1980, 1982, 1984; Lewis 1981) have observed, conditional probabilities can reflect either evidential connections or causal connections. In this example, the connection between smoking and getting cancer is merely evidential. Smoking is evidence for cancer, but it does not cause it. In deciding whether to perform an action, we consider the consequences of performing it. The consequences should be its *causal consequences*, not its evidential consequences. This suggests that a correct formulation of decision theory should replace the conditional probability **PROB**(O/A) by some kind of "causal probability". The resulting theories are called *causal decision theories.*

In formulating causal decision theory, the problem is to make sense of causal probability. It is tempting to simply replace **PROB**(O/A) by **PROB**(A causes O/A). To the best of my knowledge, no one has seriously proposed this, probably because causation is deemed too philosophically problematic to form the basis for an analysis. That is a sentiment that I share. It is also worth noting that in a world like ours, presumed to be governed by stochastic quantum mechanical processes, there may be no causes. Performing an action may "dispose" an outcome to occur by raising its probability, but it does not literally "make it happen".[1] So let us turn to other ways of making sense of causal probability.

[1] Lewis (1981) makes similar remarks about subjunctive conditionals in worlds governed by stochastic processes.

2. Subjective Probability

Most work on decision theory, causal or classical, begins with subjective probabilities, perhaps because that is the standard approach in conventional economic theory. These are *Bayesian decision theories*. There is no apparent reason why objective probabilities cannot be used instead, and that will be the course urged below. But before pursuing that, let me indicate briefly why I think subjective probability is a poor candidate for representing uncertainty in the rational deliberations of a cognizer.

Subjective probabilities are relativized to cognitive agents, and have been defined in two different ways. Sometimes the subjective probability of a proposition *P* for a person *S* is defined to be *S*'s "degree of belief" (sometimes "credence") in *P*, where that is a technical notion defined in terms of *S*'s being willing to accept bets with prescribed odds on the truth of *P*. The simplest difficulty for this definition is that it is universally acknowledged that real people do not have degrees of belief that conform to the probability calculus, and so degree of belief is not, in that sense, a probability. A person's degrees of belief are said to be *coherent* iff they conform to the probability calculus, so the observation is that real people do not have coherent degrees of belief. Subjectivists try to circumvent this by giving Dutch book arguments that purport to show that although people's degrees of belief do not conform to the probability calculus, they *should*, and a cognizer is being irrational insofar as he has degrees of belief that do not conform to the probability calculus.

On the strength of Dutch book arguments, subjectivists sometimes define the subjective probability of *P* for *S* to be the degree of belief *S* ought to have in *P*, rather than the degree of belief *S* actually has in *P*. The problem with this definition is that Dutch book arguments do not even purport to show that there is a unique degree of belief a person ought to have in *P*. Dutch book arguments purport only to establish a constraint on a rational person's overall set of degrees of belief — the degrees of belief in the set should jointly conform to the probability calculus, i.e., they should be coherent. Given that actual people do not have coherent sets of degrees of beliefs, the Dutch book argument tells us only that they should change their degrees of belief so that they become coherent. However, there are infinitely many ways to do that. The Dutch book argument gives us no guidance in how to change an incoherent set of beliefs into a coherent set of beliefs, and accordingly it gives us no reason to think there is such a thing as "the unique degree of belief a person *S* ought rationally to have in *P*". I think it must be concluded that subjective probability, defined as the degree of belief that a cognizer rationally ought to have, is a nonsensical notion.

Some philosophers, e.g., David Lewis, avoid this trap by carefully talking conditionally about how cognizers should use their actual degrees of beliefs in making decisions *if they were rational.* Here it is assumed that Dutch book arguments establish that rational cognizers will have coherent degrees of belief, and then the subsequent discussion is confined to how such cognizers can use their actual degrees of belief decision-theoretically. This makes decision theory a theory about a certain kind of ideal agent. That is not automatically objectionable, but it deserves to be emphasized how far this removes decision theory from what I want to call "the theory of rationality". I take the theory of rationality to be a theory of how, rationally, an agent should go about deciding what actions to perform at any given time. This is not a theory of real agents in the sense of being a theory about what they actually do, but it is a theory about what real agents *should* do. If it is impossible for any real agent to perform a particular computational task, then that cannot be a requirement of rationality. It is not what real agents *should do.* Properly understood, "ought" implies "can" just as much in the theory of rationality as in the theory of morality.

I take it that it is impossible (i.e., beyond the power of their limited cognitive resources) for real agents to make their degrees of belief conform to the probability calculus.[2] As such, a theory that imposes that as a requirement of rationality must be wrong. Thus we cannot save subjective probability by saying that it is only intended to make sense for rational cognizers, because rational cognizers cannot be expected to have coherent degrees of belief either. It does seem plausible to urge that rational cognizers will fix incoherences in their degrees of belief (i.e., betting behavior) when they discover them, so in this sense they may, over time, come to more closely "approximate" agents with coherent degrees of belief, but to approximate coherent degrees of belief is still not to have them, so it can never be reasonable to assume that a rational agent has coherent degrees of belief. As such, I see no way to make sense of subjective probability in a manner that makes it relevant to the rational decision making of actual cognizers.

None of this implies that one should not pursue the topic of how idealized agents should make decisions. But if the idealization is so great as to make the conclusions irrelevant to how actual agents should, rationally, make decisions, then the conclusions would seem to be of only aesthetic interest. My interests are more concrete. I want to know how the cognition of real agents should proceed when they are cognizing rationally, and for that subjective probability seems to be irrelevant. If rational cognizers must appeal to probabilities, the probabilities cannot be subjective probabilities. The only alternative is objective probabilities, so let us consider how objective probabilities work and how they should be used in decision making. In particular, how can they be used to make sense of causal probability?

---

[2] In fact, if degrees of belief are identified with degrees of justification for rational agents (and not just betting behavior), then it can be argued that it is logically impossible for an agent to systematically make its degrees of belief in newly drawn conclusions coherent. See Pollock and Cruz (1999), 106-109.

3. Nomic Probability

No doubt the strongest appeal of subjective probability has been that it seemed to provide a way of making sense of probability in the face of the failure of objective theories of probability. Traditionally, objective theories tried to define objective probability in terms of relative frequency or limits of relative frequencies, but for familiar reasons such definitions failed.[3] In the mid-twentieth century, when this dialogue was at its height, it was generally supposed that the only way to make sense of a philosophically problematic concept was by giving a definition of it in terms of simpler concepts, and that seemed to require defining objective probability in terms of relative frequency. However, by the end of the twentieth century it should be obvious to all that the program of defining complex concepts in terms of simple concepts has been a resounding failure throughout philosophy. We cannot define "red" in terms of "looks red", we cannot define mental concepts in terms of physical or behavioral concepts, we cannot define "physical object" in terms of spatio-temporal continuity, etc. The idea that concepts have to have definitions is just a bad theory of concepts. Instead, concepts can receive philosophical clarification by explaining how they are used in cognition. Thus, for example, we can solve the traditional problem of perception by describing the epistemic connection between "red" and "looks red". Something's looking red to a person gives him a defeasible reason for thinking it is red.[4] This does not derive from any deeper fact about the concepts "red" and "looks red". This relationship is a primitive constituent of our rational architecture, and philosophical clarification of the concepts can go no further than explaining how our rational architecture dictates we use the concepts in cognition.

The same lesson should be learned for probability. We are no more likely to be able to define probability in terms of simpler concepts than we are to be able to define any other complex concept in terms of simpler concepts. But perhaps we can explain objective probabilities by giving an account of how to reason about them. My (1990) purports to do just that.[5] I will give a very brief summary of some aspects of that theory, and then show how it can be used to define a kind of causal probability of use in causal decision theory.

There are two kinds of physical laws — statistical and nonstatistical. Statistical laws are probabilistic. I will call the kind of probability involved in statistical laws *nomic probability*. The

---

[3] See Pollock (1990) for a summary of those reasons.

[4] See Pollock (1986), Pollock and Cruz (1999).

[5] See also the summary of the theory of nomic probability presented in my (1992).

best way to understand nomic probability is by looking first at non-statistical laws. What distinguishes such laws from material generalizations of the form "$(\forall x)(Fx \rightarrow Gx)$" is that they are not just about actual F's. They are about "all the F's there could be", that is, they are about "physically possible F's". I call non-statistical laws *nomic generalizations*. Nomic generalizations can be expressed in English using the locution "Any F would be a G". I will symbolize this nomic generalization as "$F \Rightarrow G$". This can be roughly paraphrased as telling us that any physically possible F would be G.

I propose that we think of nomic probabilities as analogous to nomic generalizations. Just as "$F \Rightarrow G$" tells us that any physically possible F would be G, we can think of the statistical law "$\mathbf{prob}(G/F) = r$" as telling us that the proportion of physically possible F's that would be G's is r. For instance, pretend it is a law of nature that at any given time, there are exactly as many electrons as protons. Then in every physically possible world, the proportion of electrons-or-protons that are electrons is 1/2. It is then reasonable to regard the probability of a particular particle being an electron given that it is either an electron or a proton as 1/2. Of course, in the general case, the proportion of F's that are G's will vary from one possible world to another. $\mathbf{prob}(G/F)$ then "averages" these proportions across all physically possible worlds. The mathematics of this averaging process is complex, but it is discussed in detail in my (1990).

Nomic probability is illustrated by any of a number of examples that are difficult for frequency theories. For instance, consider a physical description D of a coin, and suppose there is just one coin of that description and it is never flipped. On the basis of the description D together with our knowledge of physics we might conclude that a coin of this description is a fair coin, and hence the probability of a flip of a coin of description D landing heads is 1/2. In saying this we are not talking about relative frequencies — as there are no flips of coins of description D, the relative frequency does not exist. Or suppose instead that the single coin of description D is flipped just once, landing heads, and then destroyed. In that case the relative frequency is 1, but we would still insist that the probability of a coin of that description landing heads is 1/2. The reason for the difference between the relative frequency and the probability is that the probability statement is in some sense subjunctive or counterfactual. It is not just about actual flips, but about possible flips as well. In saying that the probability is 1/2, we are saying that out of all physically possible flips of coins of description D, 1/2 of them would land heads. To illustrate nomic probability with a more realistic example, in physics we often want to talk about the probability of some event in simplified circumstances that have never occurred. For example, the typical problem given students in a quantum mechanics class is of this character. The relative frequency does not exist, but the nomic probability does and that is what the students are calculating.

The theory of nomic probability is a theory of probabilistic reasoning. No attempt is made to *define* "nomic probability" in terms of simpler concepts, because I doubt that can be done. Probabilistic reasoning has three constituents. First, there must be rules prescribing how to ascertain the numerical values of nomic probabilities on the basis of observed relative frequencies. Second, there must be "computational" principles that enable us to infer the values of some nomic probabilities from others. Finally, there must be principles enabling us to use nomic probabilities to draw conclusions about other matters.

The first element of this account will consist largely of a theory of statistical induction. Here it must be recognized that the connection between nomic probability and relative frequency is epistemological rather than definitional. The second element will consist of a calculus of nomic probabilities. The final element will be an account of how assertoric (non-probabilistic) conclusions can be inferred from premises about nomic probability. It seems clear that under some circumstances, knowing that certain probabilities are high can justify one in holding related assertoric beliefs. For example, if I want to know today's date, I normally read it off from my watch. My resulting belief is assertoric. I believe that today's date is August 14 — I do not just believe that it is probable that today's date is August 14. On the other hand, I do not believe that my watch is *always* right — just that it is extremely probable that the date displayed is today's date. So I draw an assertoric conclusion from a probabilistic premise. The epistemic rules describing when high probability can justify belief are called *acceptance rules.* The acceptance rules endorsed by the theory of nomic probability constitute the principal novelty of that theory. The other primitive assumptions about nomic probability are all of a computational nature. They concern the logical and mathematical structure of nomic probability, and amount to nothing more than an elaboration of the standard probability calculus. It is the acceptance rules that give the theory its unique flavor and comprise the main epistemological machinery making it run. The main acceptance rule employed in the theory is the following version of the *statistical syllogism*:

(SS) If F is projectible with respect to G then "$\mathbf{prob}(F/G) \geq r$" is a defeasible reason for the conditional "$Gc \rightarrow Fc$", the strength of the reason depending upon the value of r.

This is supplemented by an account of defeaters for the defeasible inference described. With the help of the computational principles comprised by the calculus of nomic probabilities, it is then shown that other kinds of defeasible inferences, such as those involved in statistical and enumerative induction, can be derived from these core defeasible inferences. In particular, the familiar projectibility constraint in induction is derived from the projectibility constraint in (SS).

To summarize, the theory of nomic probability will consist of (1) a theory of statistical induction, (2) an account of the computational principles allowing some probabilities to be derived from others, and (3) an account of acceptance rules. My (1990) provides a detailed account of all three aspects of our reasoning about nomic probabilities, and it is sketched more simply in my (1992).

4. Mixed Physical / Epistemic Probability

There is an important distinction to be made between two kinds of probabilities. Some probabilities attach to propositions. For example, subjective probabilities are probabilities of particular propositions being true. If we can make sense of them, causal probabilities are also about particular propositions. They tell us how likely it is that a certain possible consequence (a proposition) will result from performing an action under the present circumstances. The probability that a specific proposition is true is a *definite probability* (Pollock 1990). These are also called “single case probabilities”, although that is not a very good term because the propositions in question can be as general as we like. For example, we can talk about the probability that all life on earth will be destroyed if we detonate a cobalt bomb.

Nomic probabilities are not definite probabilities. Like relative frequencies, they relate classes or properties. For example, we might imagine a law telling us that the probability of an elementary particle being negatively charged is 1/2. This is not about any particular elementary particle. It is about all physically possible elementary particles, or better, about the properties of being an elementary particle and being negatively charged. Such probabilities are *indefinite probabilities*. The logical form of an indefinite probability is akin to a relative frequency. It is conditional, and we can write it with free variables. *E*.g., we can write the probability of a person with symptoms *S* having pneumonia as **prob**(*x* has pneumonia / *x* is a person with symptoms *S*). Logically, “**prob**” is a variable-binding operator, binding the variables (“*x*” in this case) in the expressions in its scope. The standard probability calculus (based on the Kolmogoroff axioms) was constructed with definite probabilities in mind. The calculus of indefinite probabilities stands to the calculus of definite probabilities as the predicate calculus stands to the propositional calculus. It is arguable that there are true principles in the calculus of indefinite probabilities that cannot be formulated in the calculus of definite probabilities. An example proposed in my (1990) is

$$\mathbf{prob}(Axy/Bxy \;\&\; y = c) = \mathbf{prob}(Axc/Bxc).$$

However, for most purposes we need only standard principles following from the Kolmogoroff axioms.

In understanding the difference between definite and indefinite probabilities, it is important not to confuse indefinite probabilities with probability distributions over random variables. The latter look superficially like indefinite probabilities, because they involve variables. However, probability distributions over random variables are *distributions* of *definite* probabilities. Using small caps to distinguish definite probabilities from indefinite probabilities, a distribution that might be written as PROB(Ax / Bx) is an assignment of a value to the *definite* probability PROB(Ac/Bc) for each choice of c in the domain of the variable. This value can be different from each choice of c. By contrast, the indefinite probability **prob**(Ax / Bx) has a single value. It is worth noting that even if the value of PROB(Ac / Bc) is the same for every choice of c, it may be different from the value of **prob**(Ax / Bx). To illustrate, consider again **prob**(x lands heads/ x is a toss of a coin of description D). General physical considerations of symmetry, etc., might convince us that a coin of this description is a fair coin, and so the indefinite probability is 0.5. However, we may also know that there has only been one coin of description D in the entire history of the universe, and there will never be another. Furthermore, we may know that it was tossed only once, and landed heads on that occasion. Then it was melted down. Given that we *know* that the one toss landed heads, the definite probability of its landing heads is 1.0. That toss is the only value for the random variable x in the distribution PROB(Ax / Bx), so for every choice of value for that variable, the definite probability is 1, but the indefinite probability remains 0.5. That is, the coin was a fair coin, and the fact that it was tossed only once and landed heads is irrelevant to whether it was a fair coin.

Understanding the relationship between definite and indefinite probabilities is of direct relevance to decision theory. The probabilities an agent learns inductively are indefinite probabilities, but the probabilities required for decision-theoretic reasoning are definite probabilities. For example, I may learn inductively that during the summer monsoon season in Tucson, the probability of late afternoon rain is 0.6. This is an indefinite probability. In deciding whether to carry an umbrella when I walk to the store, what I must know is the probability that it will rain *this afternoon*. That is a definite probability. If I have no other information, I may infer that the definite probability is 0.6. But if I also notice that the weather is clear and sunny, without any buildup of thunderheads, I may infer instead that the probability is much lower.

Our knowledge of definite probabilities is derived from our knowledge of indefinite probabilities. The kind of inference involved is called *direct inference*. The nature of direct inference is complicated by the fact that there is more than one kind of definite probability. This is illustrated by the distinction between classical decision theory and causal decision theory. Classical decision theory proceeds in terms of a “generic” definite probability PROB(O / A) that reflects both evidential and causal connections. Causal decision theory proposes to replace the appeal to PROB(O / A) by appeal to some kind of “causal” probability that is also a definite probability but focuses on causal connections. Other kinds of definite probabilities include *subjunctive probabilities* (e.g., the probability that if I were now on the moon (although I know I am not) I would be able to jump twenty feet in the air) and *objective chances* (e.g., the probability that this uranium atom will decay in the next ten minutes). I showed how to make sense of both of the latter kinds of definite probability in my (1990).

My objective here is to make sense of causal probability, but it is convenient to begin by looking at the kind of generic definite probability that is required for classical decision theory, because this has been the focus of existing theories of direct inference. These are *mixed physical/epistemic probabilities*, because they take account both of physical laws in the form of nomic probabilities and what the cognizer knows about the current situation. (As such, they are relativized to cognizers, just as subjective probabilities were.) Several theories of direct inference have been proposed in the literature on the foundations of probability theory.[6] The details of direct inference are not very well understood, although there is a consensus that the core inferences work in a way first described by Hans Reichenbach (1949). The general idea is that if we want to know $\mathsf{PROB}(Ac)$ — the definite probability of an object c having the property Ax — we identify it with the indefinite probability $\mathbf{prob}(Ax/Bx)$ where Bx is the most specific property such that (1) we know Bc to hold, and (2) we know the value of the indefinite probability. For example, in deciding how likely it is to rain this afternoon, if I know that I am in Tucson during the summer monsoon season and I know nothing else of relevance, I will identify the definite probability with the indefinite probability that it will rain in the late afternoon in Tucson during the summer monsoon season. But if I also know that the customary afternoon buildup of thunderheads is absent, and I know that the indefinite probability of its raining under those circumstances is only 0.05, then I will take the latter to be the definite probability.

In my (1983) and (1990), I proposed that Reichenbach's rules for direct inference can be reconstructed as principles for reasoning defeasibly about definite probabilities. For this we need the concept of a warranted proposition. A proposition is warranted for a cognizer just in case further reasoning from his current epistemic situation could put him in a position where he is justified in believing the proposition and in which no additional reasoning would make him unjustified. Let "$\mathbf{W}\varphi$" abbreviate "It is warranted for S that φ". Reichenbach's rules can then be reconstructed as follows:

(CDI) If A is projectible with respect to B then "$\mathbf{W}(P \leftrightarrow Ac)$ & $\mathbf{W}Bc$ & $\mathbf{prob}(Ax/Bx) = r$" is a defeasible reason for "$\mathsf{PROB}(P) = r$".

[6] Kyburg (1974, 1983), Levi (1977, 1980, 1981), Pollock (1984, 1990).

(CSD) If A is projectible with respect to C then "$\mathbf{W}Cc$ & $\mathbf{prob}(Ax\,/\,Bx) \neq \mathbf{prob}(Ax\,/\,Bx\ \&\ Cx)$" is a defeater for (CDI).

Principle CSD formulates *subproperty defeat*. It tells us that inferences from more specific properties always take precedence. For example, to reconstruct the reasoning about the weather, we note that we have two probabilities:

> **prob**(it will rain in Tucson on afternoon x / x is during the summer monsoon season) = 0.6
>
> **prob**(it will rain in Tucson on afternoon x / x is during the summer monsoon season and there is no buildup of thunderheads) = 0.05.

On the basis of (CDI), we have defeasible reasons for inferences to the conflicting conclusions:

> PROB(it will rain in Tucson this afternoon) = 0.6
>
> PROB(it will rain in Tucson this afternoon) = 0.05.

However, the second indefinite probability provides the basis for a defeater for the direct inference employing the first indefinite probability, and so we are left with the single undefeated conclusion that PROB(it will rain this afternoon) = 0.05.

(CDI) and (CSD) can be generalized in the obvious way to provide classical direct inferences to conditional probabilities:

(CDI*) If A is projectible with respect to both B and C then "$\mathbf{W}(P \leftrightarrow Ac)$ & $\mathbf{W}(Q \leftrightarrow Cc)$ & $\mathbf{W}Bc$ & $\mathbf{prob}(Ax\,/\,Bx\ \&\ Cx) = r$" is a defeasible reason for "$\mathrm{PROB}(P\,/\,Q) = r$".

(CSD*) If A is projectible with respect to D then "$\mathbf{W}Dc$ & $\mathbf{prob}(Ax\,/\,Bx\ \&\ Cx) \neq \mathbf{prob}(Ax\,/\,Bx\ \&\ Cx\ \&\ Dx)$" is a defeater for (CDI).

If we allow ourselves the full resources of first-order languages in formulating our probabilities, numerous examples show that (CDI) and (CSD) are not sufficient for a complete theory of direct inference.[7] Specifically, further defeaters are required. It is not clear exactly how to formulate them, but for present purposes (CDI) and (CSD) will probably suffice for our decision-theoretic

[7] This was first observed by Kyrburg (1974). I explored the matter further in my (1990).

reasoning.

5. Nonclassical Direct Inference

Direct inference from indefinite probabilities to definite probabilities is *classical direct inference*. In my (1984) and (1990), I introduced *nonclassical direct inference*, which is a form of defeasible inference from indefinite probabilities to indefinite probabilities. The basic principles of nonclassical direct inference are parallel to those of classical direct inference. They are formulated in terms of the concept of a subproperty. A property A is a *subproperty* of B (abbreviated "$A \preceq B$") iff $(\forall x)(Ax \rightarrow Bx)$ is logically entailed by laws of nature. So, for example, the property of being an electron is a subproperty of the property of being negatively charged. The limiting case of the subproperty relation occurs when Ax logically entails Bx. The core principles of nonclassical direct inference are then:

(DI) If A is projectible with respect to B then "$\mathbf{prob}(Ax/Bx) = r$" is a defeasible reason for "$\mathbf{prob}(Ax/Bx \mathbin{\&} Cx) = r$".

(SD) If A is projectible with respect to D then "$D \preceq C \mathbin{\&} \mathbf{prob}(Ax/Bx) \neq \mathbf{prob}(Ax/Bx \mathbin{\&} Dx)$" is a defeater for (DI).

Nonclassical direct inference amounts to a defeasible presumption that adding properties to the condition of a probability leaves the probability unchanged. This is an assumption of statistical independence.

Why should we think that (DI) is true? The simplest reason is that a slightly qualified version of (DI) is entailed by (CDI). Suppose the agent knows that Bc and Cc. Then by (CDI) it can be inferred defeasibly that $\mathrm{PROB}(Ac) = \mathbf{prob}(Ax/Bx)$, and also that $\mathrm{PROB}(Ac) = \mathbf{prob}(Ax/Bx \mathbin{\&} Cx)$. From this it follows that $\mathbf{prob}(Ax/Bx) = \mathbf{prob}(Ax/Bx \mathbin{\&} Cx)$. So we get (DI) in the special case in which the agent knows that Bc and Cc. But surely that knowledge should not make any difference to whether we can infer that $\mathbf{prob}(Ax/Bx) = \mathbf{prob}(Ax/Bx \mathbin{\&} Cx)$.

Nonclassical direct inference has somewhat the same flavor as the Laplacian principle of indifference. It amounts to the presumption that in computing probabilities, if we have no reason to think that some factor makes a difference, it is reasonable to ignore it. This is descriptive of the way we reason all the time. For example, suppose we know that quarters are "generically" fair coins. That is, the probability of a toss of a quarter landing heads is 0.5. Now consider a particular quarter. It will have many properties not shared by all quarters. For example, it was minted on a certain date. But if we have no reason to think that any of these properties make a difference, we will not hesitate to conclude that the probability of a toss of *this* quarter landing heads is also 0.5. In making this inference, we are applying nonclassical direct inference. We infer defeasibly that $\mathbf{prob}(Hx \,/\, Txy \;\&\; Qy) = \mathbf{prob}(Hx \,/\, Txy \;\&\; Qy \;\&\; y = q)$.

The preceding paragraphs provide an intuitive defense of nonclassical direct inference, but we need not rest content with that. It is proven in my (1990) that the principles (DI) and (SD) can be derived from the acceptance rule (SS) together with the computational principles comprised by the calculus of nomic probabilities. So we need not make any special assumptions in order to get the theory of nonclassical direct inference. The proof is also sketched in my (1992).

We now have two kinds of direct inference — classical and non-classical. Direct inference has traditionally been identified with classical direct inference, but I believe that it is most fundamentally non-classical direct inference. The details of classical direct inference are all reflected in non-classical direct inference. If we could identify definite probabilities with certain indefinite probabilities, we could derive the theory of classical direct inference from the theory of non-classical direct inference. This can be done by defining a variety of definite probability as a kind of degenerate indefinite probability:

> prob(P/Q) = r iff for some n, there are n-place properties R and S and objects $a_1,...,a_n$ such that $\Box(Q \leftrightarrow Sa_1...a_n)$ and $\Box[Q \rightarrow (P \leftrightarrow Ra_1...a_n)]$ and $\mathbf{prob}(Rx_1...x_n \,/\, Sx_1...x_n \;\&\; x_1 = a_1 \;\&\; ... \;\&\; x_n = a_n) = r$.

prob(P/Q) is an *objective* definite probability. It reflects the state of the world, not the state of our knowledge. This is not the same thing as mixed physical/epistemic probability, because the latter also takes account of the cognizer's epistemic state. However, if we let W be the conjunction of all warranted propositions, we can define a mixed physical/epistemic probability as follows:

> **PROB**(P) = prob(P/W)
>
> **PROB**(P / Q) = prob(P/Q&W).

Given this reduction of mixed physical/epistemic probabilities to nomic probabilities, it becomes possible to *derive* principles (CDI) and (CSD) of classical direct inference from our principles of nonclassical direct inference, and hence indirectly from (SS) and the calculus of nomic probabilities. The upshot of all this is that the theory of direct inference, both classical and nonclassical, consists of a sequence of *theorems* in the theory of nomic probability. We require no new assumptions in order to get direct inference. At the same time, we have made clear sense of mixed physical / epistemic probabilities.

6. Causal Decision Theory Again — Skyrms and Lewis

Most of the work on causal decision theory has been carried out within the framework of subjective probability. However, the basic ideas are largely independent of that, and can be reformulated in terms of objective probabilities. In this connection, let us consider Brian Skyrms' (1980, 1982, 1984) proposal. Skyrms suggests distinguishing between the *background* of an action (my terminology) and the *consequences* of an action. The background consists of states of the world that are causally independent of the performance of the action, i.e., the action does not cause or causally dispose them to occur or not occur. Let us call these *K-backgrounds*. The consequences of an action can then be evaluated against a background K by considering the probability $\mathrm{PROB}(O\,/\,A\&K)$. In computing the expected-value of an action, some parts of the background will consist of things we know to be true, but other parts of the background may be unknown to us, having only probabilities associated with them. Skyrms' suggestion is that if there is a finite set K of backgrounds that we consider possible, then in computing the expected-value we consider the probability of an outcome relative to each possible background, and weight it by the probability of that background being true. In other words, we can define causal probability as follows:

$$\text{K-PROB}_A(O) = \sum_{K\in \mathrm{K}} \mathrm{PROB}(K)\cdot\mathrm{PROB}(O\,/\,A\&K).$$

This makes causal probability the mathematical expectation of the probability of the outcome on the different possible backgrounds. It is easily verified that K-PROB_A is a probability, i.e., it satisfies the probability calculus. Then the proposal is to define expected-value in terms of $\text{K-PROB}_A(O)$ instead of $\mathrm{PROB}(O\,/\,A)$:

$$\mathbf{EV}(A) = \sum_{O\in \mathrm{O}} \mathbf{U}(O)\cdot\text{K-PROB}_A(O).$$

The K-backgrounds constitute a partition. That is, they are mutually exclusive and their disjunction is a necessary truth. Skyrms describes the K-backgrounds as maximally specific specifications of factors outside the agent's influence (at the time of the decision) which are causally relevant to the outcome of the agent's action.

Let us see how this proposal is supposed to handle the smoking gene example. Whether the person has the gene or not (G or $\sim G$) is outside his influence. If he has it, he already has it when he makes his decision whether to smoke, so that decision cannot causally influence his having or not having the gene. If G and $\sim G$ are the only elements of the background causally relevant to his getting cancer, and nothing unknown to him is relevant to his getting pleasure from smoking, then the expected-value of smoking (S) can be computed as follows:

$$\mathbf{EV}(S) = \mathbf{U}(\textit{pleasure of smoking})\cdot \textsf{K-PROB}_S(\textit{pleasure of smoking}) + \mathbf{U}(\textit{cancer})\cdot \textsf{K-PROB}_S(\textit{cancer})$$

$$= \mathbf{U}(\textit{pleasure of smoking})\cdot \textsf{PROB}(\textit{pleasure of smoking} / S) + \mathbf{U}(\textit{cancer})\cdot[\textsf{PROB}(G)\cdot \textsf{PROB}(\textit{cancer} / S\&G) + \textsf{PROB}(\sim G)\cdot \textsf{PROB}(\textit{cancer} / S\&\sim G)].$$

We have made the assumption that $\textsf{PROB}(\textit{cancer} / S\&G) = \textsf{PROB}(\textit{cancer} / G)$ and $\textsf{PROB}(\textit{cancer} / S\&\sim G) = \textsf{PROB}(\textit{cancer} / \sim G)$:

$$\mathbf{EV}(S) = \mathbf{U}(\textit{pleasure of smoking})\cdot \textsf{PROB}(\textit{pleasure of smoking} / S) + \mathbf{U}(\textit{cancer})\cdot[\textsf{PROB}(G)\cdot \textsf{PROB}(\textit{cancer} / G) + \textsf{PROB}(\sim G)\cdot \textsf{PROB}(\textit{cancer} / \sim G)].$$

Similarly, the expected-value of not smoking ($\overline{S}$) is:

$$\mathbf{EV}(\overline{S}) = \mathbf{U}(\textit{cancer})\cdot[\textsf{PROB}(G)\cdot \textsf{PROB}(\textit{cancer} / G) + \textsf{PROB}(\sim G)\cdot \textsf{PROB}(\textit{cancer} / \sim G)].$$

Thus if $\mathbf{U}(\textit{pleasure of smoking}) > 0$ and $\textsf{PROB}(\textit{pleasure of smoking} / S) > 0$, it follows that $\mathbf{EV}(S) > \mathbf{EV}(\overline{S})$. So causal decision theory recommends smoking, which is the right choice. It handles the other counterexamples to classical decision theory analogously.

David Lewis (1981) endorses a causal decision theory with the same form as Skyrms', and represents that general form as the fundamental idea behind all causal decision theories. The difference between his theory and Skyrms' is that he takes the K's to be what he calls *dependency hypotheses* — maximally specific propositions about how things the agent cares about do and do not depend causally on his present actions. Lewis proposes a narrow and a broad reading of Skyrms. On the narrow reading, the background K consists of propositions describing singular states of affairs in the world. On the broad reading, backgrounds are the same as Lewis's dependency hypotheses. Lewis observes that on the broad reading, his theory is the same as Skyrms. He goes on the argue that what he regards as the other major causal decision theories (Gibbard and Harper 1978, Sobel 1978) are also equivalent to the Skyrms/Lewis theory on at least some interpretation.

I have some difficulty understanding just what Lewis' dependency hypotheses are supposed to be. Are they supposed to be entirely relational, describing how different possible states of the world might be causally related to one another, or are they supposed to include a specification of singular states of affairs causally independent of the action? If we interpret them purely relationally, there is no apparent way to justify the calculation that is supposed to solve the smoking gene problem. This is because G and $\sim G$ would not be contained in dependency hypotheses. I presume then that dependency hypotheses must include a specification of what singular states of affairs causally independent of the action are true.

A close reading of Skyrms suggests that he did not intend Lewis' broad reading of his theory. On the contrary, K-backgrounds were supposed to consist of singular states of affairs. Lewis raises two objections to Skyrms' theory on this narrow reading. The first is that the probability of getting outcome O by performing action A depends not just on singular states of affairs, but also on laws of nature. Thus these must be included in the background. This is indeed a problem for the subjectivist (i.e., the official) version of Skyrms' theory and might reasonably be taken to motivate an expansion of the K-backgrounds to make them look more like Lewis' dependency hypothesis. However, this is not a problem when the theory is formulated in terms of objective probabilities. It is a theorem of the calculus of nomic probabilities that causal laws have probability 1. Thus there is no need for an objective causal decision theory to include them in the background. Conditionalizing on something with probability 1 cannot change the probability.

At this point the subjectivist is bound to object, "But we may be uncertain whether something is a law, so we have to attach probabilities to that and factor them into the computation of the causal probability." However, epistemic uncertainty is only represented probabilistically if you are a subjectivist. Numerous arguments throughout the epistemological literature demonstrate that epistemic degrees of justification do not conform to the probability calculus.[8] For instance, necessary truths automatically have probability 1, but it does not follow that everyone is automatically justified in believing every necessary truth. That would make mathematics trivial.

Uncertainty about the relevant laws and nomic probabilities will make one uncertain about the computation of expected-values. If we are sufficiently uncertain, e.g., we cannot even locate the relevant nomic probabilities within useful intervals, then we will not be able to draw justified conclusions about the expected-values of our actions, and so there will be no reasonable way to make a nonrandom choice between the alternatives available to us. But that seems clearly right. If I do not know enough to be able to even estimate the expected-values, then there is no way to choose. For instance, suppose I show you a button, and tell you that if you push it good things

---

[8] See Pollock and Cruz (1999), 106-109, for a summary of these objections.

might happen, but also bad things might happen, and I have no idea what the probability of either is. Can you make a rational choice between pushing or not pushing the button? Surely not. Indecision is the only possible rational attitude in this case. On the other hand, rational indecision is impossible for subjectivists (except when expected-values are tied), because for subjectivists the relevant probabilities always have to exist.

This illustrates an important difference between Bayesian decision theory and objective decision theory. In objective decision theory the expected-value of an action (relative to a person and his epistemic state) is an objective matter of fact, and decisions are made on the basis of beliefs about that objective matter of fact. It is entirely possible for the agent to be ignorant of such facts. On a subjective theory, on the other hand, the computation of expected-values is simply a matter of working out the consequences of the agent's beliefs. At least if the agent can know his own beliefs, ignorance of expected-values is impossible.

Lewis raises another objection to the narrow construal that is more telling. He observes that the *K*'s are characterized in terms of causal dependence, but cognizers may be uncertain about causal dependence and may only attach probabilities to hypotheses about causal dependence. He suggests that these probabilities should somehow enter into the computation of causal probability. Lewis maintains that his theory is not subject to this difficulty, because dependency hypotheses could not be causally dependent on the agent's actions. Skyrms (1980) takes this objection seriously, and suggests a modification of his theory that is intended to accommodate it. However, I think that this objection is indicative of a more serious objection that is telling against all current versions of causal decision theory, including Lewis's. It was observed earlier that no one has suggested defining causal probability in the most obvious way, as $\textbf{PROB}(A \text{ causes } O / A)$, presumably because no one feels sufficiently comfortable with the concept of causation to take it as a primitive building block for causal decision theory. However, all existing causal decision theories are formulated in terms of causal dependence or some similar notion, and it is hard to see how that is any clearer than "causes". In my estimation, no theory that takes causal dependence as primitive can throw adequate light on rational choice. Thus I turn in the next section to an alternative analysis of causal probability that is not subject to this objection.

7. Defining Causal Probability

My objective is to find a way of defining causal probability that does not appeal to concepts like causation or causal dependence. The basic idea behind my proposal is simple — causal probability propogates forward in time, never backward. My suggestion is that in computing the possible effects of an action, we think of the world as evolving causally over time, interject the action into the world at the appropriate point, and then propogate changes forward in time as the world continues to evolve. This way of conceptualizing the world as evolving in temporal order is precisely the same idea that underlies most current solutions to the frame problem in AI (see Shoham (1986,1987), Hanks and McDermott (1986,1987), Lifschitz (1987), Gelfond and Lifschitz (1993), Shanahan (1990,1995,1996,1997), Pollock (1998)). Those solutions are based upon the idea that given a set of deterministic causal laws, to compute the result of a sequence of actions we imagine them occurring in temporal order and propogate the changes through the world in that order. As I will define it, causal probability does the same thing probabilistically.

To make this precise, let us begin with the simplifying assumption that actions occur instantaneously. They have dates that are single instants of time. These are *point-dated* actions. I will relax this assumption later. Singular states of affairs also have dates, but I will allow them to be either time intervals or time instants (degenerate intervals). I also assume that we can assign dates to logical combinations built out of conjunctions, disjunctions, and negations of singular states of affairs. The date of a negation is the date of what it negates, the date of a conjunction is the union of the dates of the conjuncts, and the date of a disjunction is the union of the dates of the disjuncts. The date of such a combination can be a time interval with gaps. I will refer to these logical combinations of singular states of affairs as *states of affairs* (dropping "singular"). Let us say that *Q postdates P* iff every time in the date (an interval) of *P* is < every time in the date of *Q*. Let us say that *P predates Q* iff every time in the date (an interval) of *P* is ≤ every time in the date of *Q*. So if a state predates a point-dated action, the end-point of its date may be the same as the date of the action. But if it postdates the action, it occurs wholly after the date of the action.

Now suppose the world is deterministic. This means that each complete state of the world determines each subsequent state. The determination is by physical laws. Each state nomically implies subsequent states. In asking whether a possible outcome would result from a particular world state in which an action is performed, we are asking whether the outcome will be present in subsequent states. In a deterministic world, *O* will result just in case the actual state of the world up to and including the time *A* is performed includes a set *B* of singular states of affairs such that *A*&*B* nomically implies *O*. I will call *B* a *background state* for *O* relative to *A*.

If we are uncertain about the precise state of the world, then we may be uncertain about whether *O* will result. The probability that *O* will result should be identified with the probability that the state of the world at the time *A* is performed contains a background state for *O* relative to *A*. If *B* is the only background state for *O* relative to *A*, then the probability of *O* given *A* should be identified with **PROB**(*B*). If instead there are a finite number of background states $B_1,...,B_n$, then the probability of *O* given *A* should be identified with **PROB**($B_1 \vee ... \vee B_n$). Let us write this probability as $\text{C-PROB}_A(O)$.

$\text{C-PROB}_A(O)$ need not be the same as $\text{PROB}(O\,/\,A)$. The latter would be $\text{PROB}(B_1 \vee ... \vee B_n\,/\,A)$. A cannot *cause* changes to the background state, but it can be evidence regarding whether a background state is actual. This is precisely what happens in the smoking gene example. If we suppose that the gene causes cancer deterministically, then G is the sole background state and $\text{PROB}(G\,/\,A) \neq \text{PROB}(G)$. The probability $\text{C-PROB}_A(O)$ is then equal to $\text{PROB}(G)$ rather than $\text{PROB}(G\,/\,A)$. This is a *causal probability* that results from propogating the effects of actions forward in time but not backward in time. We hold the background state fixed, assigning to background states whatever probabilities they have prior to the action's being performed.

If we turn to nondeterministic worlds, the background states may no longer nomically imply the outcomes. They may only confer probabilities on the outcomes. If there were a single background B such that the only way O can only result from A by having B true, we could define

$$\text{C-PROB}_A(O) = \text{PROB}(B) \cdot \text{PROB}(O\,/\,A\&B).$$

More generally, if we could confine our attention to a finite set B of (pairwise logically disjoint) backgrounds, we could define:

$$\text{C-PROB}_A(O) = \sum_{B \in \mathrm{B}} \text{PROB}(B) \cdot \text{PROB}(O\,/\,A\&B).$$

That is, the causal probability is the mathematical expectation of the probability of the outcome on the different possible backgrounds.

To define $\text{C-PROB}_A(O)$ generally (when O postdates A and A is a point-dated action), let C be the set of all singular states of affairs and negations of singular states of affairs predating A. Define an A-*world-state* to be a maximal subset of C nomically consistent with A. I will not usually distinguish between an A-world-state and the conjunction of its members. Let W be the set of all A-world-states. Then we can define $\text{C-PROB}_A(O)$ to be the mathematical expectation of the probability of the outcome on the different possible A-world-states. If W is finite, our definition becomes:

$$\text{C-PROB}_A(O) = \sum_{W \in \mathrm{W}} \text{PROB}(W) \cdot \text{PROB}(O\,/\,A\&W).$$

Realistically, W will be infinite, in which case $\text{C-PROB}_A(O)$ must be defined using the integral definition of expected-value:

$$\textbf{C-PROB}_A(O) = \int_0^1 r \cdot \frac{\mathrm{d}}{\mathrm{d}r}\textbf{PROB}(\textbf{PROB}(O\,/\,A\&W) \le r)\,\mathrm{d}r$$

where W is a random variable ranging over members of W. However, to keep the mathematics simple, I will pretend that W is finite and use the summation version of the definition. This will make no difference to the results. Because the set W is chosen independently of O, it is trivial to verify that $\textbf{C-PROB}_A$ is a probability, i.e., that it satisfies the probability calculus.

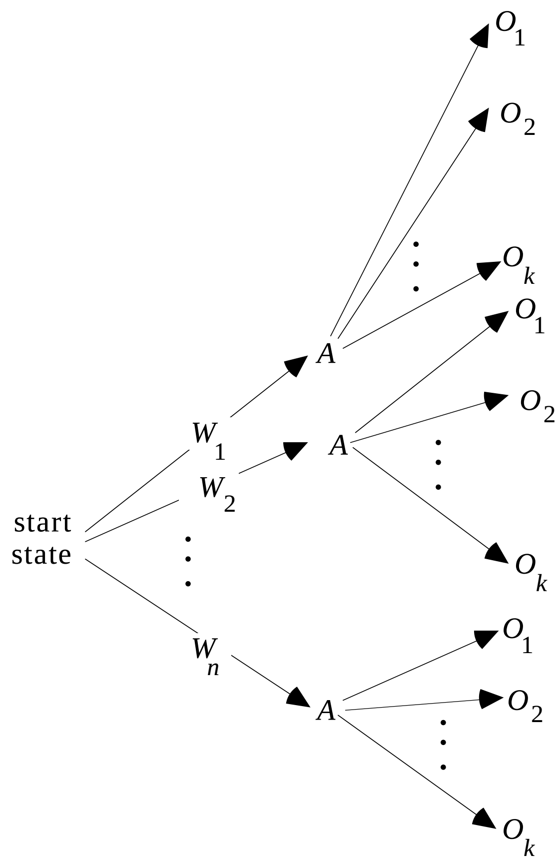

Figure 1. *S*cenarios evolving with the passage of time

The fundamental idea behind this definition of causal probability is that in computing how likely an outcome is to result from an action, we want to propogate changes forward in time rather than backward. A useful way of conceptualizing this is to think of the world as described by different *scenarios*, each consisting of some A-world-state being true, followed by the action, followed by an outcome. The scenarios can be diagrammed in the form of a tree, as in figure 1. $\textbf{C-PROB}_A(O)$ should then be of the probability of the disjunction of the scenarios terminating with O. If there were just one such scenario, the probability associated with it would be

$\textbf{PROB}(S)\cdot\textbf{PROB}(O\,/\,A\&W)$, and that would be the value of $\textbf{C-PROB}_A(O)$. The probability associated with the scenario results from propogating the probabilities of changes forward in time. We can identify $\textbf{PROB}(W)\cdot\textbf{PROB}(O\,/\,A\&W)$ with $\textbf{C-PROB}_A(O\&W)$ if we reason (1) that $\textbf{C-PROB}_A(O\&W) = \textbf{C-PROB}_A(W)\cdot\textbf{C-PROB}_A(O\,/\,W)$, (2) that $\textbf{C-PROB}_A(W) = \textbf{PROB}(W)$ because W predates A and so cannot be affected by it, and (3) that $\textbf{C-PROB}_A(O\,/\,W) = \textbf{PROB}(O\,/\,A\&W)$ because W includes everything that is relevant to whether O will result from performing A.

Suppose there are n possible scenarios, associated with the A-world-states $W_1,\ldots,W_n$. Then the disjunction of the scenarios is $(W_1\&A\&O) \vee\ldots\vee (W_n\&A\&O)$, which is equivalent to $(W_1 \vee\ldots\vee W_n)\&A\&O$. We can think of this as a single scenario with a disjunctive background state and identify $\textbf{C-PROB}_A(O)$ with the probability of this scenario. The different W_i's are logically disjoint, so the probability associated with this scenario can be computed as follows:

$$\textbf{C-PROB}_A(O) = \textbf{C-PROB}_A((W_1 \vee\ldots\vee W_n)\&A\&O)$$

$$= \textbf{C-PROB}_A(W_1 \vee\ldots\vee W_n)\cdot\textbf{C-PROB}_A(O\,/\,W_1 \vee\ldots\vee W_n)$$

$$\begin{aligned}= \textbf{C-PROB}_A(W_1 \vee\ldots\vee W_n) &\\ \cdot[\textbf{C-PROB}_A(O\,/\,W_1)\cdot\textbf{C-PROB}_A(W_1\,/\,W_1 \vee\ldots\vee W_n) &\\ + \ldots + \textbf{C-PROB}_A(O\,/\,W_n)\cdot\textbf{C-PROB}_A(W_n\,/\,W_1 \vee\ldots\vee W_n)]&\end{aligned}$$

$$= \textbf{C-PROB}_A(O\,/\,W_1)\cdot\textbf{C-PROB}_A(W_1) + \ldots + \textbf{C-PROB}_A(O\,/\,W_n)\cdot\textbf{C-PROB}_A(W_n)$$

$$= \textbf{PROB}(W_1)\cdot\textbf{C-PROB}_A(O\,/\,W_1) + \ldots + \textbf{PROB}(W_n)\cdot\textbf{C-PROB}_A(O\,/\,W_n)$$

$$= \textbf{PROB}(W_1)\cdot\textbf{PROB}(O\,/\,A\&W_1) + \ldots + \textbf{PROB}(W_n)\cdot\textbf{PROB}(O\,/\,A\&W_n).$$

In other words, $\textbf{C-PROB}_A(O)$ is the probability associated with the disjunctive scenario, and that in turn is the sum of the probabilities associated with the individual scenarios.

If causal probability is to be useful, there must be efficient ways of computing it. If we had to compute $\textbf{C-PROB}_A(O)$ by actually performing the summation (or integration) involved in the definition, the task would be formidable. Fortunately, this computation can be simplified considerably. Recall that C is the set of “constituents” of A-world-states. Let us say that a subset S of C *shadows* A with respect to O iff (1) S is nomically consistent with A, (2) for every $W \in$ W and any S^{**}, if $S \subseteq S^{**} \subseteq W$ then $\textbf{PROB}(O\,/\,A\&S^{**}) = \textbf{PROB}(O\,/\,A\&S)$, and (3) there is no proper subset S^*

of S such that for every $W \in \mathrm{W}$ and any S^{**}, if $S^* \subseteq S^{**} \subseteq W$ then $\mathbf{PROB}(O / A \& S^{**}) = \mathbf{PROB}(O / A \& S)$. The shadows are minimal descriptions of all aspects of the A-world-state relevant to the evaluation of the probability of O. Let S be the set of all shadows. Shadows can be constructed by starting from members of W and then removing elements that do not affect the probability of O. It follows that every A-world-state W contains a shadow S such that $\mathbf{PROB}(O / A \& W) = \mathbf{PROB}(O / A \& S)$.

Let C* be the set of all members of C occurring in one or more of the shadows. Define a *background* to be a maximal subset of C* nomically consistent with A. Let B be the set of all backgrounds. The backgrounds form a partition. That is, they are pairwise logically disjoint and the disjunction of all of them is a (nomically) necessary truth. Now suppose $B \in \mathrm{B}$ is a background and $W \in \mathrm{W}$ and $B \subseteq W$. W contains a shadow S such that $\mathbf{PROB}(O / A \& W) = \mathbf{PROB}(O / A \& S)$, and the shadow consists of members of C*, so $S \subseteq B$, and hence by the definition of "shadow", $\mathbf{PROB}(O / A \& B) = \mathbf{PROB}(O / A \& S)$. Thus $\mathbf{PROB}(O / A \& W) = \mathbf{PROB}(O / A \& B)$. Now we can prove a central theorem in the theory of causal probability:

Theorem 1: $\mathbf{C\text{-}PROB}_A(O) = \sum_{B \in \mathrm{B}} \mathbf{PROB}(B) \cdot \mathbf{PROB}(O / A \& B)$.

Proof: For $B \in \mathrm{B}$, let $W(B) = \{W \mid W \in \mathrm{W} \;\&\; B \subseteq W\}$. B is equivalent to the disjunction of the members of $W(B)$. Then

$$\mathbf{C\text{-}PROB}_A(O) = \sum_{W \in \mathrm{W}} \mathbf{PROB}(W) \cdot \mathbf{PROB}(O / A \& W)$$

$$= \sum_{B \in \mathrm{B}} \sum_{W \in W(B)} \mathbf{PROB}(W) \cdot \mathbf{PROB}(O / A \& W)$$

$$= \sum_{B \in \mathrm{B}} \sum_{W \in W(B)} \mathbf{PROB}(W) \cdot \mathbf{PROB}(O / A \& B)$$

$$= \sum_{B \in \mathrm{B}} \mathbf{PROB}(O / A \& B) \cdot \sum_{W \in W(B)} \mathbf{PROB}(W)$$

$$= \sum_{B \in \mathrm{B}} \mathbf{PROB}(O / A \& B) \cdot \mathbf{PROB}(B). \quad \blacksquare$$

W is immense (in fact, infinite), but B may be very small. In the smoking gene example, if we suppose that the only part of an S-world-state that makes any difference to the probability of getting cancer is G or $\sim G$, it follows that $\mathrm{S} = \{\{G\},\{\sim G\}\}$, and so $\mathrm{B} = \{\{G\},\{\sim G\}\}$, and hence

$$\textbf{C-PROB}_S(cancer) = \textbf{PROB}(G)\cdot\textbf{PROB}(cancer\,/\,S\&G) + \textbf{PROB}(\sim G)\cdot\textbf{PROB}(cancer\,/\,S\&\sim G).$$

Of course, realistically, other elements of S-world-states will also be statistically relevant, e.g., whether one's parents had the smoking gene. However, the effect of one's parents having the smoking gene is "screened off" by knowing whether one has the gene oneself, i.e., if you know whether you have the smoking gene, the additional knowledge of whether your parents had it does not effect the probability of getting cancer. So the set of shadows, and hence the set of backgrounds, remains unchanged.

Normally, shadows will be more numerous than in the smoking gene example. However, the shadows may not all be relevant. The need for causal probabilities only arises when the action is statistically relevant to some of the backgrounds. If the backgrounds are all statistically independent of the action, then the causal probability is the same as the mixed physical / epistemic probability:

Theorem 2: If for each $B\in \mathrm{B}$, $\textbf{PROB}(B\,/\,A) = \textbf{PROB}(B)$, then $\textbf{C-PROB}_A(O) = \textbf{PROB}(O\,/\,A)$.

More generally, the action may be statistically relevant to just a few constituents of the backgrounds. Then we can often make use of the following theorem:

Theorem 3: If $\mathrm{C}_0 \subseteq \mathrm{C}^*$, let B_0 be the set of all maximal subsets of C_0 nomically consistent with A, and let B^* be the set of all maximal subsets of $\mathrm{C}^* - \mathrm{C}_0$ nomically consistent with A. If for every $B_0\in \mathrm{B}_0$ and $B^*\in \mathrm{B}^*$, $\textbf{PROB}(B^*\,/\,B_0\&A) = \textbf{PROB}(B^*\,/\,B_0)$, then

$$\textbf{C-PROB}_A(O) = \sum_{B\in \mathrm{B}_0} \textbf{PROB}(B_0)\cdot\textbf{PROB}(O\,/\,A\&B_0).$$

Proof: The backgrounds B are just the conjunctions (unions) of a $B_0\in \mathrm{B}_0$ and a $B^*\in \mathrm{B}^*$, and the disjunction of the members of B^* is necessary, so

$$\textbf{C-PROB}_A(O) = \sum_{B\in \mathrm{B}} \textbf{PROB}(B)\cdot\textbf{PROB}(O\,/\,A\&B)$$

$$= \sum_{B_0\in \mathrm{B}_0} \sum_{B^*\in \mathrm{B}^*} \textbf{PROB}(B_0\&B^*)\cdot\textbf{PROB}(O\,/\,A\&B_0\&B^*)$$

$$= \sum_{B_0 \in \mathrm{B}_0} \sum_{B^* \in \mathrm{B}^*} \textbf{PROB}(B_0) \cdot \textbf{PROB}(B^* / B_0) \cdot \textbf{PROB}(O / A \& B_0 \& B^*)$$

$$= \sum_{B_0 \in \mathrm{B}_0} \textbf{PROB}(B_0) \cdot \sum_{B^* \in \mathrm{B}^*} \textbf{PROB}(B^* / B_0) \cdot \textbf{PROB}(O / A \& B_0 \& B^*)$$

$$= \sum_{B_0 \in \mathrm{B}_0} \textbf{PROB}(B_0) \cdot \sum_{B^* \in \mathrm{B}^*} \textbf{PROB}(B^* / B_0 \& A) \cdot \textbf{PROB}(O / A \& B_0 \& B^*)$$

$$= \sum_{B \in \mathrm{B}_0} \textbf{PROB}(B_0) \cdot \textbf{PROB}(O / A \& B_0). \ \blacksquare$$

So if there is a subset C_0 of constituents of backgrounds relative to which all other combinations of constituents are statistically independent of A, then we can compute causal probabilities by making reference only to backgrounds built out of the members of C_0. For example, suppose there are two constituents of backgrounds that are statistically relevant to getting cancer — having the smoking gene, and having been raised on a nuclear waste dump (N). Then $\mathrm{B} = \{\{G,N\},\{G,\sim N\},\{\sim N,G\},\{\sim N,\sim G\}\}$. However, S is not statistically relevant to whether one was raised on a nuclear waste dump, even given that one does or does not have the smoking gene:

$$\textbf{PROB}(N / S \& G) = \textbf{PROB}(N / G)$$
$$\textbf{PROB}(N / S \& \sim G) = \textbf{PROB}(N / \sim G)$$
$$\textbf{PROB}(\sim N / S \& G) = \textbf{PROB}(\sim N / G)$$
$$\textbf{PROB}(\sim N / S \& \sim G) = \textbf{PROB}(\sim N / \sim G)$$

So we can let $\mathrm{C}_0 = \{\{G\},\{\sim G\}\}$, and once more compute $\textbf{C-PROB}_S(\textit{cancer})$ by reference to the small set of backgrounds $\mathrm{B}_0 = \{\{G\},\{\sim G\}\}$.

The upshot of these results is that causal probabilities will usually be computable by performing manageably small sums. In cases in which actions are statistically relevant to their backgrounds, **C-PROB**'s may be significantly easier to compute than **PROB**'s. **C-PROB**'s can be computed recursively by propogating probabilities forwards through scenarios. But if a later state can affect the **PROB** of an earlier state, then **PROB**'s cannot similarly be computed recursively. For practical purposes, **C-PROB**'s are simpler than **PROB**'s. This suggests that instead of expressing theorem 2 by saying that causal probabilities usually behave classically, it might be better to say that classical probabilities usually behave causally.

Thus far, $\text{C-PROB}_A(O)$ has been defined for all states of affairs postdating A. It will be convenient to define $\text{C-PROB}_A(O)$ for a broader class of states of affairs, including states of affairs that do not postdate A. If O predates A we can stipulate:

$$\text{C-PROB}_A(O) = \text{PROB}(O).$$

If O_1 postdates A and O_2 predates A, then we will further stipulate that

$$\text{C-PROB}_A(O_1 \,\&\, O_2) = \text{PROB}(O_1)\cdot\text{C-PROB}_A(O_2 / O_1).$$

(However, conditional causal probability will not be defined until section nine.) We are making the simplifying assumption that actions occur instantaneously, and so their dates are time points rather than intervals. If a state of affairs neither predates A nor postdates A then its date must be an interval (possibly with gaps) with the date of A lying within the interval. I assume that such a state of affairs can be split into a "first part" predating A and a "second part" postdating A, and then the state of affairs can be represented as the conjunction of these two parts. This has the consequence that $\text{C-PROB}_A(O)$ is defined for all states of affairs O.

We can construct a version of causal decision theory by defining expected-values in terms of C-PROB:

$$\mathbf{EV}(A) = \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\text{C-PROB}_A(O).$$

I will call this *T-causal decision theory* because of its reliance on temporal ordering rather than causal dependence.

8. C-PROB_A and K-PROB_A

T-causal decision theory handles the counterexamples to classical decision theory in essentially the same way other causal decision theories do, but it defines causal probability without appeal to causation or causal dependence. It seems to me that the appeal to the evolution of scenarios in temporal order is a more obvious diagnosis of the counterexamples than is the appeal to causal dependence. It resolves the counterexamples in an intuitively congenial way, without appealing to anything more problematic than temporal ordering and the fact that causation propogates forwards in time.

Without an analysis, the concept of causal dependence is sufficiently unclear that the behavior of a concept of causal probability defined in terms of it is not clear either. However, if we confine our attention to point-dated actions and singular states of affairs postdating A and make two plausible assumptions about causal dependence, it follows that T-causal decision theory is equivalent to Skyrms' theory (on the narrow construal). First, we need to assume that if P predates A, then P is causally independent of A. I take it that this is obvious and uncontroversial. It has the consequence that the elements of a background W in W are causally independent of A. Because K is a complete specification of causally independent states of affairs, it follows that if $K \in$ K then there is a background $W(K)$ in W such that $W(K) \subseteq K$. K will also contain many states of affairs postdating A. Most of them will be statistically independent of A in the sense that, if K_0 is the set of them, then $\mathbf{PROB}(K_0 / A\&(K–K_0)) = \mathbf{PROB}(K_0 / (K–K_0))$. It then follows as in theorem 3 that omitting them from K will not affect the calculation of $\mathbf{K\text{-}PROB}_A(O)$. For the remaining elements of K that postdate A, A is statistically relevant to them but they are causally independent of A. My second assumption is that this is only possible if the elements of K and A have a common cause. To be causally relevant to A, that common cause must lie in the part of K that predates A, i.e., $W(K)$. So the precise assumption I will make is that $\mathbf{PROB}(O / A\&K) = \mathbf{PROB}(O / A\&W(K))$. It is to be emphasized that this is a considerable precization of a rather vague assumption about causal relevance. Let K* = $\{K – W(K) \mid K \in$ K$\}$. For any $W \in$ W, W is equivalent to the disjunction of all $W\&K^*$ for $K^* \in$ K*. So

$$\sum_{K^* \in \mathrm{K}^*} \mathbf{PROB}(W\&K^*) = \mathbf{PROB}(W).$$

Then we can compute:

$$\begin{aligned}
\mathbf{K\text{-}PROB}_A(O) &= \sum_{K \in \mathrm{K}} \mathbf{PROB}(O / A\&K)\cdot\mathbf{PROB}(K) \\
&= \sum_{K \in \mathrm{K}} \mathbf{PROB}(O / A\& W(K))\cdot\mathbf{PROB}(K) \\
&= \sum_{W \in \mathrm{W}} \sum_{K^* \in \mathrm{K}^*} \mathbf{PROB}(O / A\& W(K^*\&W))\cdot\mathbf{PROB}(K^*\&W) \\
&= \sum_{W \in \mathrm{W}} \sum_{K^* \in \mathrm{K}^*} \mathbf{PROB}(O / A\& W)\cdot\mathbf{PROB}(K^*\&W) \\
&= \sum_{W \in \mathrm{W}} \mathbf{PROB}(O / A\&W)\cdot\sum_{K^* \in \mathrm{K}^*} \mathbf{PROB}(K^*\&W)
\end{aligned}$$

$$= \sum_{W \in \mathrm{W}} \textbf{PROB}(O / A \& W) \cdot \textbf{PROB}(W)$$

$$= \textbf{C-PROB}_A(O).$$

Thus if we make these two assumptions about causal dependence, $\textbf{C-PROB}_A(O) = \textbf{K-PROB}_A(O)$, and hence T-causal decision theory is equivalent to Skyrms' theory. However, T-causal decision theory has the advantage that causal probability is defined without reference to causation or causal dependence. Should the second assumption be false, then I suggest that $\textbf{C-PROB}_A$ is the more obvious choice for causal probability and handles the counterexamples to classical decision theory more tidily than do other causal decision theories.

9. Conditional Policies and Conditional Causal Probability

Decision theory has usually focussed on choosing between alternative actions available to us *here and now*. A generalization of this problem is important in some contexts. We sometimes make *conditional decisions* about what to do if some condition P turns out to be true. For instance, I might deliberate about what route to take to my destination if I encounter road construction on my normal route. Where P predates A, *doing A if P* is a *conditional policy*. Conditional decisions are choices between conditional policies. We can regard noncausal decision theory as telling us to make such conditional decisions on the basis of the expected-values of the conditional policies. The simplest way to handle the conditional policy in noncausal decision theory is to take it to be equivalent to the disjunction ($\sim P \vee P \& A$). Then if expected-values are computed classically, it follows that the expected-value of the conditional policy A *if* P is just the expected-value of A discounted by the probability of P plus the expected-value of doing nothing discounted by the probability of $\sim P$:

$$\mathbf{EV}(A \textit{ if } P)$$
$$= \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \textbf{PROB}(O / \sim P \vee P \& A).$$

It is a theorem of the probability calculus that

$$\textbf{PROB}(O / \sim P \vee P \& A) = \textbf{PROB}(P / \sim P \vee A) \cdot \textbf{PROB}(O / A \& P) + \textbf{PROB}(\sim P / \sim P \vee A) \cdot \textbf{PROB}(O / \sim P).$$

Thus

$$\mathbf{EV}(A\ if\ P)$$
$$= \sum_{O\in \mathcal{O}} \mathbf{U}(O)\cdot\left[\textsf{PROB}(P / {\sim}P\vee A)\cdot\textsf{PROB}(O / A\&P) + \textsf{PROB}({\sim}P / {\sim}P\vee A)\cdot\textsf{PROB}(O / {\sim}P)\right]$$
$$= \textsf{PROB}(P / {\sim}P\vee A)\cdot\sum_{O\in \mathcal{O}}\mathbf{U}(O)\cdot\textsf{PROB}(O / A\&P)$$
$$+\ \textsf{PROB}({\sim}P / {\sim}P\vee A)\cdot\sum_{O\in \mathcal{O}}\mathbf{U}(O)\cdot\textsf{PROB}(O / {\sim}P).$$

In causal decision we would similarly like to be able to define

$$\mathbf{EV}(A\ if\ P) = \sum_{O\in \mathcal{O}}\mathbf{U}(O)\cdot\textsf{C-PROB}_{A\ if\ P}(O).$$
$$\mathbf{EV}(A\ if\ P\ /\ Q) = \sum_{O\in \mathcal{O}}\mathbf{U}(O)\cdot\textsf{C-PROB}_{A\ if\ P}(O / Q).$$

For this we must define the causal probability of O conditional on execution of the conditional policy. We might propose:

$$\textsf{C-PROB}_{A\ if\ P}(O) = \textsf{C-PROB}_{A\ if\ P}(P)\cdot\textsf{C-PROB}_{A}(O / P) + \textsf{C-PROB}_{A\ if\ P}({\sim}P)\cdot\textsf{C-PROB}_{nil}(O/{\sim}P).$$

But this does not constitute a definition, because $\textsf{C-PROB}_{A\ if\ P}$ occurs on the right side of the equation. However, P predates A, so we should have $\textsf{C-PROB}_{A\ if\ P}(P) = \textsf{PROB}(P)$ and $\textsf{C-PROB}_{A\ if\ P}({\sim}P) = \textsf{PROB}({\sim}P)$. This allows us to turn the preceding principle into a definition:

$$\textsf{C-PROB}_{A\ if\ P}(O) = \textsf{PROB}(P)\cdot\textsf{C-PROB}_{A}(O / P) + \textsf{PROB}({\sim}P)\cdot\textsf{C-PROB}_{nil}(O/{\sim}P).$$

$$\textsf{C-PROB}_{A\ if\ P}(O / Q) = \textsf{PROB}(P / Q)\cdot\textsf{C-PROB}_{A}(O / P\&Q) + \textsf{PROB}({\sim}P / Q)\cdot\textsf{C-PROB}_{nil}(O/{\sim}P\&Q).$$

This definition proceeds in terms of conditional probabilities, and we have yet to define those. The standard definition would be:

$$\textsf{C-PROB}_{A}(O / P) = \textsf{C-PROB}_{A}(O\&P) / \textsf{C-PROB}_{A}(P).$$

Unfortunately, when P predates A, we defined $\textsf{C-PROB}_{A}(O\&P)$ in terms of $\textsf{C-PROB}_{A}(O / P)$, so we must find an independent definition for the latter. Two possibilities may occur to us regarding how to do that:

(1) $\textbf{C-PROB}_A(O\,/\,P) = \sum_{W \in \mathbf{W}} \textbf{PROB}(W) \cdot \textbf{PROB}(O\,/\,A\&W\&P)$.

(2) $\textbf{C-PROB}_A(O\,/\,P) = \sum_{W \in \mathbf{W}} \textbf{PROB}(W\,/\,P) \cdot \textbf{PROB}(O\,/\,A\&W\&P)$.

The issue is whether we should conditionalize the probability of the backgrounds on *P*. We can answer this by modifying the smoking gene example. To keep the mathematics simple, suppose that smoking is neither pleasurable nor unpleasant. From that perspective there is no reason to prefer either smoking or not smoking to the other alternative. As before, suppose the smoking gene is rare, but wanting to smoke makes it more probable that one has the smoking gene. However, the significance of the smoking gene is different than it was before. For normal people (those lacking the smoking gene), smoking tends (weakly) to cause lung cancer, however the smoking gene protects people from that. Then if you know you have the smoking gene and you desire to smoke, you might as well do it. Both classical decision theory and causal decision theory agree on this prescription. And if you know that you lack the smoking gene, then both classical decision theory and causal decision theory agree that you should not smoke.

Now let us add a twist to the example. Suppose that for most people, the smell of tobacco smoke is an acquired taste. When they first smell tobacco smoke, it repels them. However, for some people, when they first smell tobacco smoke they experience an almost overpowering urge to smoke. The latter trait is quite rare, but it is an infallible indicator of the presence of the smoking gene. Suppose you have never smelled tobacco smoke. You are now deliberating on whether to smoke if, when you first smell tobacco smoke, you experience this overpowering urge to smoke. That indicates that you have the smoking gene, in which case smoking will not hurt you. So you might as well smoke. Classical decision theory yields the right prescription. What about causal decision theory? We have

$$\begin{aligned}
&\mathbf{EV}(\textit{smoke if have-urge}) \\
&= \mathbf{U}(\textit{lung cancer}) \cdot \textbf{C-PROB}_{\textit{smoke if have-urge}}(\textit{lung cancer}) \\
&= \mathbf{U}(\textit{lung cancer}) \cdot [\textbf{PROB}(\textit{have-urge}) \cdot \textbf{C-PROB}_{\textit{smoke}}(\textit{lung cancer}\,/\,\textit{have-urge}) \\
&\qquad + \textbf{PROB}(\sim\textit{have-urge}) \cdot \textbf{C-PROB}_{\textit{nil}}(\textit{lung cancer}\,/\,\sim\textit{have-urge})].
\end{aligned}$$

Similarly

$$\mathbf{EV}(\textit{not-smoke if have-urge})$$

$$= \mathbf{U}(\textit{lung cancer})\cdot[\textsf{PROB}(\textit{have-urge})\cdot\textsf{C-PROB}_{\textit{not-smoke}}(\textit{lung cancer} / \textit{have-urge}) + \textsf{PROB}(\sim\textit{have-urge})\cdot\textsf{C-PROB}_{\textit{nil}}(\textit{lung cancer} / \sim\textit{have-urge})].$$

Smoking is permissible iff

$$\mathbf{EV}(\textit{smoke if have-urge}) \geq \mathbf{EV}(\textit{not-smoke if have-urge}).$$

As $\mathbf{U}(\textit{lung cancer}) < 0$, this holds iff

$$\textsf{C-PROB}_{\textit{smoke}}(\textit{lung cancer} / \textit{have-urge}) \leq \textsf{C-PROB}_{\textit{not-smoke}}(\textit{lung cancer} / \textit{have-urge}).$$

Smoking is permissible if we define conditional causal probabilities as in (2). But if we define them as in (1), causal decision theory will proscribe smoking, because at this point, when you do not yet know whether you will have the overpowering urge to smoke, it is very improbable that you have the smoking gene and hence somewhat probable that smoking will cause lung cancer. Thus conditional decisions require conditional causal probability to be defined as in (2). We can conceptualize this in terms of scenarios by replacing “start state” by *P* in figure 1.

When *P* predates *A*, our official definition is:

$$\textsf{C-PROB}_A(O / P) = \sum_{W \in \mathrm{W}} \textsf{PROB}(W / P)\cdot\textsf{PROB}(O / A\&W\&P).$$

This has the consequence that *P* functions informationally while *A* functions causally. That is, *P* can have backward ramifications, influencing the probability of backgrounds, but *A* can only influence the probabilities of future events. This definition can be recast in terms of backgrounds, just as in theorem 1:

Theorem 4: If *P* predates *A* then where B is the set of backgrounds for *O* relative to *A* that are consistent with *P*:

$$\text{C-PROB}_A(O/P) = \sum_{B\in \mathrm{B}} \text{PROB}(B/P)\cdot\text{PROB}(O/A\&B\&P).$$

The proof is analogous to that of theorem 1. For any P predating A, let $W(P)$ be the set of all W in W consistent with P. Note that P is equivalent to the disjunction of members of $W(P)$, and if $W \notin W(P)$ then $\text{PROB}(W/P) = 0$. Thus

$$\text{C-PROB}_A(O/P)$$

$$= \sum_{W\in \mathrm{W}} \text{PROB}(W/P)\cdot\text{PROB}(O/A\&W\&P)$$

$$= \sum_{W\in W(P)} \text{PROB}(W/P)\cdot\text{PROB}(O/A\&W\&P)$$

$$= \sum_{W\in W(P)} \text{PROB}(W)\cdot\text{PROB}(O/A\&W\&P)/\text{PROB}(P)$$

$$= \sum_{B\in \mathrm{B}}\sum_{W\in W(B\&P)} \text{PROB}(W)\cdot\text{PROB}(O/A\&W\&P)/\text{PROB}(P)$$

$$= \sum_{B\in \mathrm{B}}\sum_{W\in W(B\&P)} \text{PROB}(W)\cdot\text{PROB}(O/A\&B)/\text{PROB}(P)$$

$$= \sum_{B\in \mathrm{B}} (\text{PROB}(O/A\&B)/\text{PROB}(P))\cdot\sum_{W\in W(B\&P)} \text{PROB}(W)$$

$$= \sum_{B\in \mathrm{B}} (\text{PROB}(O/A\&B)/\text{PROB}(P))\cdot\text{PROB}(B\&P)$$

$$= \sum_{B\in \mathrm{B}} \text{PROB}(B/P)\cdot\text{PROB}(O/A\&B\&P). \quad \blacksquare$$

From this we get a simple theorem that will be repeatedly useful:

Theorem 5: If B is a background for A relative to O and P predates A then $\text{C-PROB}_A(O/B\&P)$ =

$\textsf{PROB}(O / A\&B\&P)$.

Proof: $\textsf{PROB}(B / B\&P) = 1$, and for any other background B^*, $\textsf{PROB}(B^* / B\&P) = 0$. ■

We also get theorems analogous to theorems 2 and 3:

Theorem 6: If P predates A and for every background B for A relative to O that is consistent with P, $\textsf{PROB}(B / A\&P) = \textsf{PROB}(B / P)$, then $\textsf{C-PROB}_A(O / P) = \textsf{PROB}(O / A\&P)$.

Theorem 7: If $C_0 \subseteq C^*$, let B_0 be the set of all maximal subsets of C_0 nomically consistent with $A\&P$, and let B^* be the set of all maximal subsets of $C^* - C_0$ nomically consistent with $A\&P$. If for every $B_0 \in B_0$ and $B^* \in B^*$, $\textsf{PROB}(B^* / B_0\&A\&P) = \textsf{PROB}(B^* / B_0\&P)$, then

$$\textsf{C-PROB}_A(O / P) = \sum_{B \in B_0} \textsf{PROB}(B_0 / P) \cdot \textsf{PROB}(O / A\&B_0\&P).$$

By virtue of these theorems, in computing conditional causal probabilities we can usually restrict our attention to very small backgrounds.

Once conditional probabilities are defined as above for the case in which P predates A, non-conditional causal probabilities are defined in general (for point-dated actions), and so for all other cases we can stipulate conventionally that:

$$\textsf{C-PROB}_A(O / P) = \textsf{C-PROB}_A(O\&P) / \textsf{C-PROB}_A(P).$$

Just as for nonconditional probabilities, the conditional causal probabilities of states predating A behave classically:

Theorem 8: If P and Q predate A, $\textsf{C-PROB}_A(Q / P) = \textsf{PROB}(Q / P)$.

Proof: $\textsf{C-PROB}_A(Q / P) = \textsf{C-PROB}_A(Q\&P) / \textsf{C-PROB}_A(P) = \textsf{PROB}(Q\&P) / \textsf{PROB}(P) = \textsf{PROB}(Q / P)$. ■

It follows from theorem 8 that conditional policies are probabilitistically irrelevant to states predating the action (as was presupposed by our definition of $\textsf{C-PROB}_{A \text{ } if \text{ } P}$):

Theorem 9: If Q predates A, $\textsf{C-PROB}_{A \text{ } if \text{ } P}(Q) = \textsf{PROB}(Q)$.

Proof: $\text{C-PROB}_{A \text{ if } P}(Q) = \text{PROB}(P)\cdot\text{C-PROB}_A(Q/P) + \text{PROB}(\sim P)\cdot\text{PROB}(Q/\sim P) = \text{PROB}(P)\cdot\text{PROB}(Q/P) + \text{PROB}(\sim P)\cdot\text{PROB}(Q/\sim P) = \text{PROB}(Q)$. ■

We say that a case is *classical* iff for every background B for A relative to O, $\text{PROB}(B/A\&P) = \text{PROB}(B/P)$. By theorems 2 and 6, in classical cases $\text{C-PROB}_A(O) = \text{PROB}(O/A)$ and $\text{C-PROB}_A(O/P) = \text{PROB}(O/A\&P)$. However, despite the fact that the definition of $\text{C-PROB}_{A \text{ if } P}(O)$ was motivated by the classical calculation of $\text{PROB}(O/\sim P\vee A)$, even in classical cases it will not usually be true that $\text{C-PROB}_{A \text{ if } P}(O) = \text{PROB}(O/\sim P\vee A)$. The classical theorem tells us that

$$\text{PROB}(O/\sim P\vee A) = \text{PROB}(P/\sim P\vee A)\cdot\text{PROB}(O/A\&P) + \text{PROB}(\sim P/\sim P\vee A)\cdot\text{PROB}(O/\sim P).$$

If we define:

$$\text{PROB}_{A \text{ if } P}(O) = \text{PROB}(P)\cdot\text{PROB}(O/A\&P) + \text{PROB}(\sim P)\cdot\text{PROB}(O/\sim P)$$

then it is easily proven that in classical cases, $\text{C-PROB}_{A \text{ if } P}(O) = \text{PROB}_{A \text{ if } P}(O)$. However, there is no guarantee that $\text{PROB}_{A \text{ if } P}(O) = \text{PROB}(O/\sim P\vee A)$ in classical cases. This is because $\text{PROB}(P/\sim P\vee A)$ will normally be different from $\text{PROB}(P)$. We can compute:

$$\text{PROB}(P/\sim P\vee A) = \text{PROB}(P/A)\cdot\text{PROB}(A/\sim P\vee A) + \text{PROB}(P/\sim P\&\sim A)\cdot\text{PROB}(\sim A/\sim P\vee A)$$

$$= \text{PROB}(P/A)\cdot\text{PROB}(A/\sim P\vee A).$$

P predates A, so we would normally expect that $\text{PROB}(P/A) = \text{PROB}(P)$. However, we would also normally expect that $\text{PROB}(A/\sim P\vee A) < 1$, in which case it follows that $\text{PROB}(P/\sim P\vee A) < \text{PROB}(P)$.

What this actually shows is that even in classical cases it is not reasonable to identify the conditional policy A *if* P with the disjunction $(\sim P\vee A)$. P is serving as a *trigger* for A, and so what should be relevant is $\text{PROB}(P)$ rather than $\text{PROB}(P/\sim P\vee A)$. In other words, classically, the expected-value of the conditional policy should be defined in terms of $\text{PROB}_{A \text{ if } P}(O)$ rather than $\text{PROB}(O/\sim P\vee A)$.

It is important to distinguish between the expected-value of a conditional policy and a conditional expected-value. The latter can be defined as follows:

$$\mathbf{EV}(A\,/\,P) = \sum_{O\in \mathcal{O}} \mathbf{U}(O)\cdot \textsf{C-PROB}_A(O\,/\,P).$$

This is the expected-value of the action given the assumption that P is true. **Expected-value**(A *if* P), on the other hand, is the expected-value of doing A if P and doing nothing otherwise. The expected-value of a conditional policy is related to conditional expected-values as follows:

Theorem 10: $\mathbf{EV}(A \textit{ if } P) = \textsf{PROB}(P)\cdot\mathbf{EV}(A\,/\,P) + \textsf{PROB}(\sim P)\cdot\mathbf{EV}(nil\,/\,{\sim}P)$.

Proof:

$$\mathbf{EV}(A \textit{ if } P)$$

$$= \sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot[\textsf{PROB}(P)\cdot\textsf{C-PROB}_{try\text{-}A}(O\,/\,A) + \textsf{PROB}(\sim P)\cdot\textsf{C-PROB}_{nil}(O\,/\,{\sim}P)]$$

$$= \textsf{PROB}(P)\cdot\sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot\textsf{C-PROB}_{try\text{-}A}(O\,/\,A) + \textsf{PROB}(\sim P)\cdot\sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot\textsf{C-PROB}_{nil}(O\,/\,{\sim}P)$$

$$= \textsf{PROB}(P)\cdot\mathbf{EV}(A\,/\,P) + \textsf{PROB}(\sim P)\cdot\mathbf{EV}(nil\,/\,{\sim}P).\ \blacksquare$$

The following theorem will be useful later:

Theorem 11: $\mathbf{EV}(nil \textit{ if } P) = \mathbf{EV}(nil)$

Proof:

$$\mathbf{EV}(nil \textit{ if } P)$$

$$= \textsf{PROB}(P)\cdot\mathbf{EV}(nil\,/\,P) + \textsf{PROB}(\sim P)\cdot\mathbf{EV}(nil\,/\,{\sim}P)$$

$$= \sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot[\textsf{PROB}(P)\cdot\textsf{C-PROB}_{nil}(O\,/\,P) + \textsf{PROB}(\sim P)\cdot\textsf{C-PROB}_{nil}(O\,/\,{\sim}P)]$$

$$= \sum_{O\in\mathcal{O}} \mathbf{U}(O)\cdot\sum_{B\in \mathcal{B}} [\textsf{PROB}(P)\cdot\textsf{PROB}(B\,/\,P)\cdot\textsf{PROB}(O\,/\,nil\&B\&P)$$

$$+ \textbf{PROB}(\sim P)\cdot\textbf{PROB}(B / \sim P)\cdot\textbf{PROB}(O / nil\&B\&\sim P)]$$

$$= \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\sum_{B\in \mathrm{B}} [\textbf{PROB}(P)\cdot\textbf{PROB}(B / P)\cdot\textbf{PROB}(O / B\&P)$$

$$+ \textbf{PROB}(\sim P)\cdot\textbf{PROB}(B / \sim P)\cdot\textbf{PROB}(O / B\&\sim P)]$$

$$= \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\sum_{B\in \mathrm{B}} [\textbf{PROB}(O\&B\&P) + \textbf{PROB}(O\&B\&\sim P)]$$

$$= \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\sum_{B\in \mathrm{B}} \textbf{PROB}(O\&B)$$

$$= \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\sum_{B\in \mathrm{B}}\textbf{PROB}(B)\cdot\textbf{PROB}(O / nil\&B)$$

$$= \sum_{O\in \mathrm{O}}\mathbf{U}(O)\cdot\textbf{C-PROB}_{nil}(O)$$

$$= \mathbf{EV}(nil). \quad \blacksquare$$

Decision theory normally concerns itself with expected-values. However, the expected-value of an action is defined to be the expected-value of "the world" when the action is performed. This includes values that would have been achieved with or without the action. For some purposes it is more useful to talk about the *marginal expected-value*, which is the difference between the expected-value of the action and the expected-value of doing nothing. The marginal expected-value of an action measures how much value the action can be expected to *add* to the world:

$$\mathbf{MEV}(A) = \mathbf{EV}(A) - \mathbf{EV}(nil).$$

We can define conditional marginal expected-values and the marginal expected-values of conditional policies analogously:

$$\mathbf{MEV}(A / P) = \mathbf{EV}(A / P) - \mathbf{EV}(nil / P).$$

$$\mathbf{MEV}(A \text{ } if \text{ } P) = \mathbf{EV}(A \text{ } if \text{ } P) - \mathbf{EV}(nil \text{ } if \text{ } P).$$

Note that by theorem 11 we could just as well have defined:

$$\mathbf{MEV}(A \textit{ if } P) = \mathbf{EV}(A \textit{ if } P) - \mathbf{EV}(nil).$$

There is a simple relationship between the marginal expected-value of a conditional policy and the conditional marginal expected-value:

Theorem 12: $\mathbf{MEV}(A \textit{ if } P) = \text{PROB}(P)\cdot\mathbf{MEV}(A / P)$

Proof:

$$\mathbf{MEV}(A \textit{ if } P)$$

$$= \mathbf{EV}(A \textit{ if } P) - \mathbf{EV}(nil \textit{ if } P)$$

$$= \text{PROB}(P)\cdot\mathbf{EV}(A / P) + \text{PROB}(\sim P)\cdot\mathbf{EV}(nil / \sim P)$$

$$- \text{PROB}(P)\cdot\mathbf{EV}(nil / P) - \text{PROB}(\sim P)\cdot\mathbf{EV}(nil / \sim P)$$

$$= \text{PROB}(P)\cdot\mathbf{EV}(A / P) - \text{PROB}(P)\cdot\mathbf{EV}(nil / P)$$

$$= \text{PROB}(P)\cdot[\mathbf{EV}(A / P) - \mathbf{EV}(nil / P)]$$

$$= \text{PROB}(P)\cdot\mathbf{MEV}(A / P). \quad \blacksquare$$

10. Linear Policies

Thus far we have considered how to define the causal probability of an outcome given a single action. In decision-theoretic contexts we will often want to consider what is apt to happen if we perform several actions sequentially. Furthermore, when we relax the assumption that actions occur instantaneously, we will find that it is often desirable to decompose a single action into temporal parts and treat it as a sequence of actions. Let *linear policies* be sequences of (for now, point-dated) actions in which each action postdates its predecessor. First consider a linear policy consisting of two actions A_1,A_2. Computing the causal probability of an outcome is complicated by the fact that some constituents of the background of the second action may postdate the first action, and the first action can affect the probabilities of those constituents. So let B be the set of backgrounds for A_1 conjoined with those parts of the backgrounds of A_2 that predate A_1, and let B* consist of the remainders of the backgrounds for A_2. The members of B* postdate A_1. If we conceptualize the world as evolving in temporal order as in figure 2 and assume that O postdates A_2, the probability associated with a scenario should be

$$\textbf{PROB}(B_i)\cdot\textbf{PROB}(B^*_j / A_1\&B_i)\cdot\textbf{PROB}(O / A_1\&A_2\&B_i\&B^*_j).$$

Then we can define:

$$\textbf{C-PROB}_{A_1,A_2}(O) = \sum_{B\in \mathrm{B}} \textbf{PROB}(B) \cdot \sum_{B^*\in \mathrm{B}^*} \textbf{PROB}(B^* / A_1\&B)\cdot\textbf{PROB}(O / A_1\&A_2\&B\&B^*)$$

$$= \sum_{B\in \mathrm{B}} \textbf{PROB}(B)\cdot\textbf{C-PROB}_{A_2}(O / A_1\&B).$$

This definition can be generalized recursively to arbitrary sequences (for $k > 1$) of point-dated actions postdated by O:

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O) = \sum_{B\in \mathrm{B}} \textbf{PROB}(B)\cdot\textbf{C-PROB}_{A_2,\ldots,A_k}(O / A_1\&B).$$

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O / Q) = \sum_{B\in \mathrm{B}} \textbf{PROB}(B / Q)\cdot\textbf{C-PROB}_{A_2,\ldots,A_k}(O / A_1\&B\&Q).$$

Again, this calculation is what we get from propogating probabilities through scenarios in temporal order.

Recalling that if O predates A_2 then $\textbf{C-PROB}_{A_2}(O / A_1\&B) = \textbf{PROB}(O / A_1\&B)$, it follows that $\textbf{C-PROB}_{A_1,A_2}(O) = \textbf{C-PROB}_{A_1}(O)$. More generally:

Theorem 13: If O postdates A_i and O predates A_{i+1} then

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O / Q) = \textbf{C-PROB}_{A_1,\ldots,A_i}(O / Q).$$

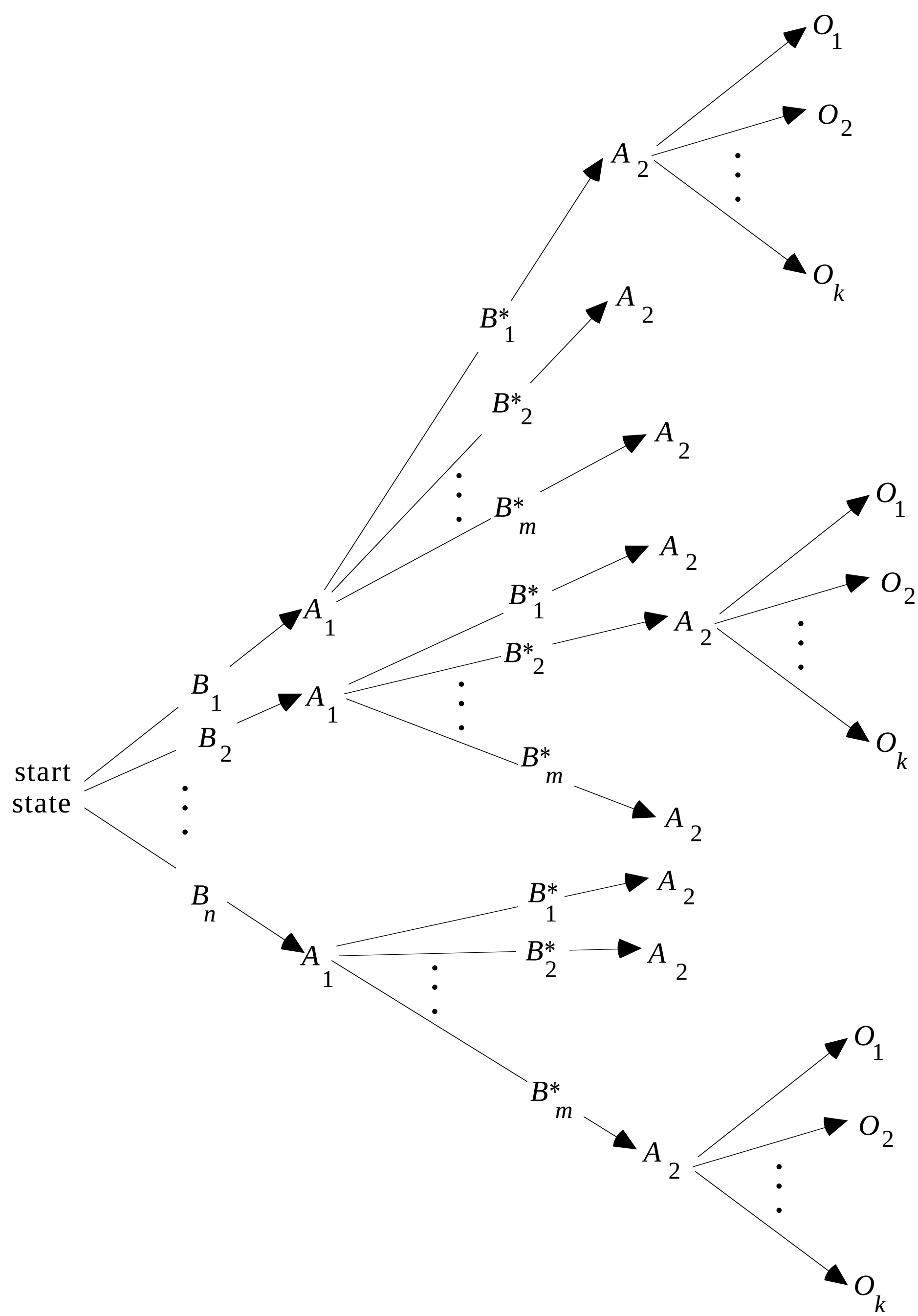

Figure 2. Scenarios with two actions

Two other useful theorems are analogous to theorem 5:

Theorem 14: If B is a background for O given A_1 then $\textsf{C-PROB}_{A_1,A_2}(O/B) = \textsf{C-PROB}_{A_2}(O/A_1\&B)$.

Theorem 15: If B is a background for B^* given A_1 and $B\&B^*$ is a background for O given A_2 then

$$\text{C-PROB}_{A_1,A_2}(O\,/\,B\&B^*) = \text{PROB}(O\,/\,A_1\&A_2\&B\&B^*).$$

Given the policy A_1,A_2, consider the scenario $B \rightarrow A_1 \rightarrow B^* \rightarrow A_2$. Let us assume that O postdates A_2. Then by theorems 14 and 15 we can compute:

$$\begin{aligned}\text{C-PROB}_{A_1,A_2}(B\&B^*\&O) &= \text{PROB}(B)\cdot\text{C-PROB}_{A_1,A_2}(B^*\&O\,/\,B)\\ &= \text{PROB}(B)\cdot\text{C-PROB}_{A_1}(B^*/\,B)\cdot\text{C-PROB}_{A_1,A_2}(O\,/\,B\&B^*)\\ &= \text{PROB}(B)\cdot\text{PROB}(B^*/\,A_1\&B)\cdot\text{PROB}(O\,/\,A_1\&A_2\&B\&B^*)\end{aligned}$$

Thus

$$\text{C-PROB}_{A_1,A_2}(O) = \sum_{B\in \mathrm{B}} \sum_{B^*\in \mathrm{B}^*} \text{C-PROB}_{A_1,A_2}(B\&B^*\&O)$$

$$= \sum_{B\in \mathrm{B}} \sum_{B^*\in \mathrm{B}^*} \text{C-PROB}_{A_1,A_2}(B\&B^*)\cdot\text{C-PROB}_{A_1,A_2}(O\,/\,B\&B^*)$$

$$= \sum_{B\in \mathrm{B}} \sum_{B^*\in \mathrm{B}^*} \text{C-PROB}_{A_1,A_2}(B\&B^*)\cdot\text{PROB}(O\,/\,A_1\&A_2\&B\&B^*).$$

$\text{C-PROB}_{A_1,A_2}(B\&B^*)$ is the probability of the scenario given the policy. Let us identify the scenario with the conjunction ($B\&B^*$). Then where SC is the set of scenarios, this can be expressed equivalently as:

$$\text{C-PROB}_{A_1,A_2}(O) = \sum_{S\in \mathrm{SC}} \text{C-PROB}_{A_1,A_2}(S)\cdot\text{C-PROB}_{A_1,A_2}(O\,/\,S).$$

Similar reasoning establishes this in general:

Theorem 16: If O postdates A_k and SC is the set of scenarios for the linear policy $A_1,\ldots,A_k$ relative to O then

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O/Q) = \sum_{S\in \mathrm{SC}} \textbf{C-PROB}_{A_1,\ldots,A_k}(S/Q)\cdot\textbf{C-PROB}_{A_1,\ldots,A_k}(O/S\&Q).$$

More generally,

Theorem 17: If O postdates A_i and O predates A_{i+1} and SC is the set of scenarios for the linear policy $A_1,\ldots,A_i$ relative to O then

$$\textbf{C-PROB}_{A_1,\ldots,A_k}(O/Q) = \sum_{S\in \mathrm{SC}} \textbf{C-PROB}_{A_1,\ldots,A_i}(S/Q)\cdot\textbf{C-PROB}_{A_1,\ldots,A_i}(O/S\&Q).$$

11. Conditional Linear Policies

As thus-far construed, linear policies are simple sequences of actions. In decision-theoretic planning, the plans have a more complex structure. They can be viewed as *conditional linear policies*, which are sequences of conditional policies rather than sequences of actions. Let A_1 *if* $C_1,\ldots,A_k$ *if* C_k be the policy *do* A_1 *if* C_1, *then do* A_2 *if* C_2, *then* ... . As for simple conditional policies, the probabilities of the B_i's and the probabilities of the outcomes must be made conditional on the C_i's. If C_i is false, the rest of the policy will still be executed. So on analogy to simple conditional policies, for $k > 1$ the causal probability can be defined recursively as:

$$\begin{aligned}&\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O)\\ &\qquad = \textbf{PROB}(C_1)\cdot\sum_{B\in \mathrm{B}} \textbf{PROB}(B/C_1)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/A_1\&B\&C_1)\\ &\qquad\quad + \textbf{PROB}(\sim C_1)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/\sim C_1).\end{aligned}$$

$$\begin{aligned}&\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O/Q)\\ &\qquad = \textbf{PROB}(C_1/Q)\cdot\sum_{B\in \mathrm{B}} \textbf{PROB}(B/C_1\&Q)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/A_1\&B\&C_1\&Q)\\ &\qquad\quad + \textbf{PROB}(\sim C_1/Q)\cdot\textbf{C-PROB}_{A_2\ if\ C_2,\ldots,A_k\ if\ C_k}(O/\sim C_1\&Q).\end{aligned}$$

We can then define the expected-value of a conditional linear policy in terms of these causal probabilities:

$$\mathbf{EV}(A_1\ if\ C_1,\ldots,A_k\ if\ C_k) = \sum_{O\in \mathrm{O}} \mathbf{U}(O)\cdot\textbf{C-PROB}_{A_1\ if\ C_1,\ldots,A_k\ if\ C_k}(O).$$

A scenario for a conditional linear policy looks like this:

$$(\sim)C_1 \rightarrow B_1 \rightarrow (A_1) \rightarrow (\sim)C_2 \rightarrow B_2 \rightarrow (A_2) \rightarrow \ldots \rightarrow (\sim)C_k \rightarrow B_k \rightarrow (A_k)$$

where each tilde can be present or absent, and if it is absent on C_i then A_i is included in the scenario. Otherwise A_i is are not included. So a scenario is characterized by the set of unnegated C_i's. For example, the following is a scenario:

$$C_1 \rightarrow B_1 \rightarrow A_1 \rightarrow \sim C_2 \rightarrow B_2 \rightarrow C_3 \rightarrow B_3 \rightarrow A_3$$

Given a scenario S, let C_S be the conjunction of the C_i's, $\sim C_i$'s, and B_i's in the scenario, and let A_S be the conjunction of the actions in the scenario. Defining the probability of a scenario as $\textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(C_S)$, we again get:

Theorem 18: If O postdates A_k and SC is the set of scenarios for the policy A_1 *if* $C_1,\ldots,A_k$ *if* C_k relative to O then

$$\textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(O / Q)$$
$$= \sum_{S \in \mathrm{SC}} \textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S / Q) \cdot \textbf{PROB}(O / C_S \& A_S \& Q).$$

More generally:

Theorem 19: If O postdates A_i and O predates A_{i+1} and SC is the set of scenarios for the policy A_1 *if* $C_1,\ldots,A_i$ *if* C_i relative to O then:

$$\textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(O / Q)$$
$$= \sum_{S \in \mathrm{SC}} \textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_i \textit{ if } C_i}(S / Q) \cdot \textbf{PROB}(O / C_S \& A_S \& Q).$$

We can define the expected-value of a scenario in the obvious way:

$$\mathbf{EV}(S) = \sum_{O \in \mathrm{O}} \mathbf{U}(O) \cdot \textbf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(O / C_S).$$

It then follows that:

Theorem 20: If SC is the set of scenarios for the policy A_1 *if* $C_1,\ldots,A_k$ *if* C_k then:

$$\mathbf{EV}(A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k) = \sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot\mathbf{EV}(S).$$

We get an analogous theorem about marginal expected-values:

Theorem 21: If SC is the set of scenarios for the policy A_1 *if* $C_1,\ldots,A_k$ *if* C_k then:

$$\mathbf{MEV}(A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k) = \sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot\mathbf{MEV}(S).$$

Proof: The disjunction of the scenarios is equivalent to the policy, so

$$\sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S) = 1.$$

Then we can compute:

$$\mathbf{MEV}(A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k) = \mathbf{EV}(A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k) - \mathbf{EV}(nil)$$

$$= \sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot\mathbf{EV}(S)$$

$$- \mathbf{EV}(nil)\cdot\sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)$$

$$= \sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot[\mathbf{EV}(S) - \mathbf{EV}(nil)]$$

$$= \sum_{S\in \mathrm{SC}} \textsf{C-PROB}_{A_1 \textit{ if } C_1,\ldots,A_k \textit{ if } C_k}(S)\cdot\mathbf{MEV}(S). \blacksquare$$

Thus the marginal expected-value of a conditional linear policy is a weighted average of the marginal expected-values of its scenarios.

The next theorem tells us that the marginal expected-value of a scenario is the sum of the marginal expected-values of its actions in the context of the scenario:

Theorem 22: If S is a scenario and $A_1,...,A_n$ are the actions it prescribes listed in temporal order then:

$$\mathbf{MEV}(S) = \sum_{1\leq i\leq n}\mathbf{MEV}(A_i / A_1 \,\&...\&\, A_{i-1} \,\&\, C_S).$$

Proof:

Where S is a scenario, let $A_1,...,A_n$ be the actions it prescribes, listed in temporal order. Then by theorem 5, $\mathbf{EV}(S) = \mathbf{EV}(A_1,...,A_n / C_S)$. Note that

$$\mathbf{MEV}(A_i / A_1 \,\&...\&\, A_{i-1} \,\&\, C_S)$$

$$= \mathbf{EV}(A_i / A_1 \,\&...\&\, A_{i-1} \,\&\, C_S) - \mathbf{EV}(nil / A_1 \,\&...\&\, A_{i-1} \,\&\, C_S)$$

$$= \mathbf{EV}(A_1 \,\&...\&\, A_i / C_S) - \mathbf{EV}(A_1 \,\&...\&\, A_{i-1} / C_S).$$

It follows that

$$\sum_{1\leq i\leq k}\mathbf{MEV}(A_i / A_1 \,\&...\&\, A_{i-1} \,\&\, C_S)$$

$$= \mathbf{EV}(A_1 \&...\&\, C_k \,\&\, C_S) - \mathbf{EV}(A_1 \&...\&\, C_{k-1} \,\&\, C_S)$$

$$+ \mathbf{EV}(A_1 \,\&...\&\, C_{k-1} \,\&\, C_S) - \mathbf{EV}(A_1 \&...\&\, C_{k-2} \,\&\, C_S)$$

$$+ ... + \mathbf{EV}(A_1 \,\&\, C_S) - \mathbf{EV}(\, C_S)$$

$$= \mathbf{EV}(A_1 \&...\&\, C_k \,\&\, C_S) - \mathbf{EV}(C_S)$$

$$= \mathbf{MEV}(S). \;\blacksquare$$

Theorems 19 and 20 together tell us how to compute the marginal expected-value of a linear policy (and hence a decision-theoretic plan) in terms of the marginal expected-values of its actions in the possible scenarios of the policy. Unfortunately, the computation this prescribes will often be very difficult. The problem is that it requires us to compute marginal expected-values for every scenario separately, and compute the marginal expected-value of the policy or plan as a weighted average of the marginal expected-values of the scenarios. There can be a very large number of scenarios.

Sometimes it will be possible to ignore the individual scenarios and compute marginal expected-values more directly. For this purpose, let us define a different kind of conditional marginal expected-value:

$$\mathbf{MEV}(A_i \textit{ if } C_i // A_1 \textit{ if } C_1,...,A_{i-1} \textit{ if } C_{i-1}) = \mathbf{EV}(A_1 \textit{ if } C_1,...,A_i \textit{ if } C_i) - \mathbf{EV}(A_1 \textit{ if } C_1,...,A_{i-1} \textit{ if } C_{i-1}).$$

Then the following theorem tells us that we can compute $\mathbf{MEV}(A_1 \textit{ if } C_1,...,A_k \textit{ if } C_k)$ by summing these conditional marginal expected-values:

Theorem 23: $\mathbf{MEV}(A_1 \textit{ if } C_1,...,A_k \textit{ if } C_k) = \sum_{1 \leq i \leq k} \mathbf{MEV}(A_i \textit{ if } C_i // A_1 \textit{ if } C_1,...,A_{i-1} \textit{ if } C_{i-1})$.

The proof is analogous to that of theorem 22. If we can compute the conditional marginal expected-values $\mathbf{MEV}(A_i \textit{ if } C_i // A_1 \textit{ if } C_1,...,A_{i-1} \textit{ if } C_{i-1})$, this theorem makes it easy to compute the marginal expected-value of the policy. Unfortunately, the conditional marginal expected-values are generally hard to compute without appealing to scenarios (in which case, nothing is saved).

12. Actions with Temporal Duration

Now let us relax the assumption that actions occur instantaneously. I will allow the date of an action to be an interval, possibly with gaps. If the backgrounds predate the action and the outcome postdates it, I presume that $\textbf{C-PROB}_A(O)$ can be defined just as for point-dated actions. The definitions for linear policies need not be changed either as long as the elements of the backgrounds either predate the actions or fall between them. However, if there is temporal overlap between the action and the backgrounds or outcome, matters become much more complicated. Let us begin by supposing the backgrounds predate the action but the date of the action overlaps the date of the outcome. We can distinguish two cases.The first occurs when O has a date $[t_0,t_3]$ and A has a date $[t_1,t_2]$, where $t_0 < t_1 < t_2 < t_3$. We can divide O into three parts, one predaing A, one with the same date as A, and one postdating A. We can write these as $O[t_0,t_1]$,

$O(t_1,t_2]$, and $O(t_2,t_3]$. O is equivalent to the conjunction of these three states of affairs, so we ought to have:

$$\textbf{C-PROB}_A(O) = \textbf{C-PROB}_A(O[t_0,t_1])\cdot\textbf{C-PROB}_A(O(t_1,t_2]\,/\,O[t_0,t_1])\cdot\textbf{C-PROB}_A(O(t_2,t_3]\,/\,O[t_0,t_2]).$$

Because $O[t_0,t_1]$ predates A, $\textbf{C-PROB}_A(O[t_0,t_1]) = \textbf{PROB}(O[t_0,t_1])$. As $O(t_2,t_3]$ postdates A, we can compute $\textbf{C-PROB}_A(O(t_2,t_3]/O[t_0,t_2])$ as before. But as yet we have no definition of $\textbf{C-PROB}_A(O(t_1,t_2]/O[t_0,t_1])$. The second problem is essentially the converse. We can have a state of affairs O with date $[t_1,t_2]$ and an action A with date $[t_0,t_3]$, where $t_0 < t_1 < t_2 < t_3$. In this case we can divide the action into three parts $A[t_0,t_1)$, $A[t_1,t_2]$, and $A(t_2,t_3]$ where O posdates the first part and predates the third part, but O has the same date as the middle part. The action can then be treated as a linear policy consisting of three separate actions. Then we know how to compute $\textbf{C-PROB}_A(O)$ if we can compute $\textbf{C-PROB}_{A[t_1,t_2]}(O)$. The remaining problem is how to define $\textbf{C-PROB}_A(O)$ when O and A have the same date. I do not have a complete solution to this problem, but I do have some ideas that may eventually point in the direction of a solution.

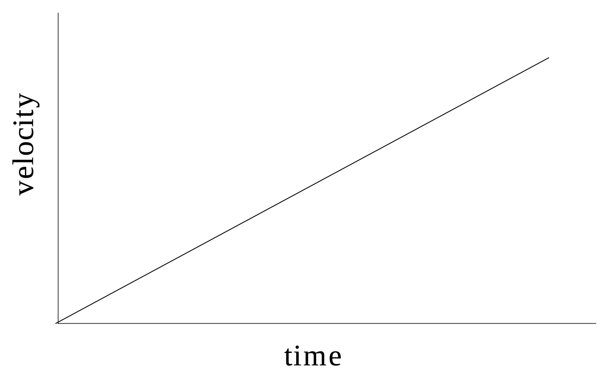

Figure 3. Velocity vs. time

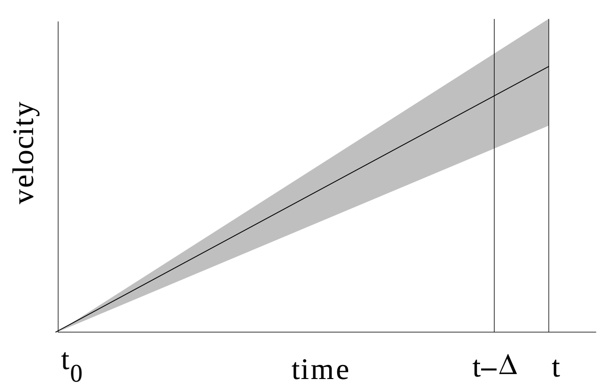

Figure 4. Probability envelope

What makes this problem hard is that the relatinship between A and O can reflect continuous causal processes. For example, suppose $A[t_0,t]$ consists of applying a constant force to a particle initially at rest and $O[t_0,t]$ describes the velocity of the particle at the different times throughout the interval. If no other forces act on the particle, and the world is Newtonian, then the velocities are described by the graph in figure 3. However, in the real world we never know precisely what other forces are acting on a particle. Suppose that for any time between t_0,t_1 we know with some prescribed probability that the acceleration on the particle is $\alpha \pm \varepsilon$. Then there is an envelope of velocities within which the actual velocity can be expected to fall with that probability. This is diagrammed in figure 4. If the force was applied at discrete instants followed by discrete changes in the velocity, we could compute the causal probabilities as before. But in reality the changes are continuous. This suggests, however, that we might be able to characterize the continuous case as the limit to which the discrete cases go as the intervals between the changes go to zero. This might lead to a characterization of the derivative of the probability, and then the probability could be computed by integrating the derivative over the interval $[t_0,t_1]$:

$$\textbf{C-PROB}_{A[t_0,t_1]}(O[t_0,t_1]) = \int_{t_0}^{t_1} \frac{d}{dt}\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t])\, \mathrm{d}t$$

and figuring out a way of computing the derivative. To compute the derivative, note that where Δ is some small number, we can compute:

$$\textbf{C-PROB}_{A}(O[t_0,t+\Delta])$$

$$= \textbf{C-PROB}_{A}(O[t_0,t])\cdot\textbf{C-PROB}_{A}(O(t,t+\Delta] \,/\, O[t_0,t])$$

$$= \textbf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O[t_0,t])\cdot\textbf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O(t,t+\Delta] \,/\, O[t_0,t]).$$

$O[t_0,t]$ predates $A(t,t+\Delta]$, so

$$\textbf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O[t_0,t]) = \textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]).$$

Thus

$$\textsf{C-PROB}_{A[t_0,t+\Delta]}(O[t_0,t+\Delta]) - \textsf{C-PROB}_{A[t_0,t]}(O[t_0,t])$$

$$= \textsf{C-PROB}_{A[t_0,t]}(O[t_0,t])\cdot[\textsf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O(t,t+\Delta]\,/\,O[t_0,t]) - 1].$$

$$= - \textsf{C-PROB}_{A[t_0,t]}(O[t_0,t])\cdot\textsf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(\sim O(t,t+\Delta]\,/\,O[t_0,t]).$$

Hence

$$\frac{d}{dt}\textsf{C-PROB}_{A[t_0,t]}(O[t_0,t])$$
$$= - \textsf{C-PROB}_{A[t_0,t]}(O[t_0,t])\cdot\lim_{\Delta\to 0}\ \frac{\textsf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(\sim O(t,t+\Delta]\,/\,O[t_0,t])}{\Delta}$$

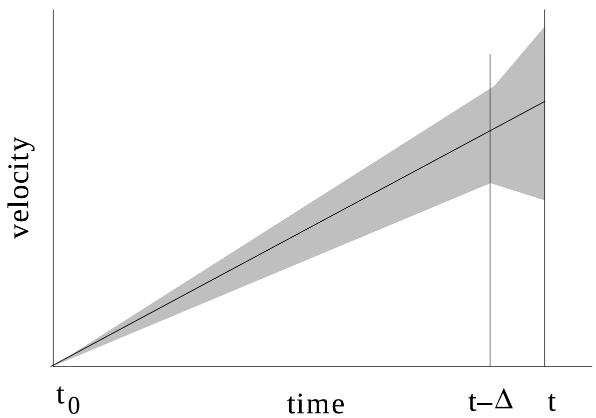

Figure 5. Broadening the envelope

In evaluating a probability of the form $\textsf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O(t,t+\Delta]\,/\,O[t_0,t])$, it usually seems reasonable to conclude that the contribution of $A(t,t+\Delta]$ becomes vanishingly small as Δ goes to zero. In the example of the accelerating particle, if we remove the assumption that the force acts continuously on the particle between t–Δ and t, that has the effect of broadening that part of the envelope, as in figure 5. As Δ goes to zero, the difference between the two envelopes becomes vanishingly small. In other words,

$$\lim_{\Delta\to 0}\textsf{C-PROB}_{A[t_0,t],A(t,t+\Delta]}(O(t,t+\Delta]\,/\,O[t_0,t])$$

$$= \lim_{\Delta\to 0}\textsf{C-PROB}_{A[t_0,t]}(O(t,t+\Delta]\,/\,O[t_0,t]).$$

So we can conclude:

$$\frac{d}{dt}\,\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t])$$

$$= -\,\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t])\cdot\lim_{\Delta\to 0}\frac{\textbf{C-PROB}_{A[t_0,t]}(\sim O(t,t+\Delta]\,/\,O[t_0,t])}{\Delta}$$

In the probability $\textbf{C-PROB}_{A[t_0,t]}(O(t,t+\Delta]\,/\,O[t_0,t])]$, the action predates the outcome and so is well-defined.

Unfortunately, "$\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t])$" occurs in the expression of the derivative, so what this gives us is a differential equation rather than a simple derivative. Sometimes it will have a desirable solution. For example, suppose the envelope is constructed in such a way that we can conclude that for some constant β,

$$\textbf{C-PROB}_{A[t_0,t]}(\sim O(t,t+\Delta]\,/\,O[t_0,t]) = \beta\cdot\Delta.$$

Then the differential equation we get is

$$\frac{d}{dt}\,\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]) = -\beta\cdot\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]).$$

This equation can be solved to yield:

$$\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]) = e^{-\beta(t-t_0)}.$$

This gives us a nice expression of the probability. However, it is also possible for the differential equation we get to have multiple solutions. For instance, if we are able to conclude instead that

$$\textbf{C-PROB}_{A[t_0,t]}(\sim O(t,t+\Delta]\,/\,O[t_0,t]) = \frac{\beta}{\varepsilon(t-t_0)}$$

then our differential equation becomes

$$\frac{d}{dt}\,\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]) = -\frac{\beta}{\varepsilon(t-t_0)}\,\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]).$$

This equation has multiple solutions of the form

$$\textbf{C-PROB}_{A[t_0,t]}(O[t_0,t]) = -\frac{\beta}{n\varepsilon}(t-t_0)^n.$$

This yields quite different probabilities for different choices of n. Thus in this case this approach is not sufficient to uniquely determine a value for $\textsf{C-PROB}_A(O)$.

In those cases in which the preceding approach generates a differential equation with a unique solution, we can take it as defining $\textsf{C-PROB}_A(O)$ when A and O have the same interval date. But as we have seen, this will not always generate a definition. In those cases, it is unclear how to proceed. This is a matter for future research.

If we relax the assumption that the backgrounds do not temporally overlap the actions, things get even more complicated, but the general ideas remain the same. I will not pursue the details.

13. Computing Causal Probability

Decision theory is a theory about how cognizers should, rationally, direct their activities. Causal decision theory tells them to use causal probabilities in their deliberations, and to do that they must have beliefs about causal probabilities. How can such probabilities be computed?

Causal probabilities are defined in terms of mixed physical/epistemic probabilities, which are in turn inferred by direct inference from nomic probabilities. To apply the definitions directly a cognizer would have to know all the relevant nomic probabilities and compute all the relevant physical/epistemic probabilities. Real cognizers will fall far short of this ideal. They have limited knowledge of nomic probabilities, and correspondingly limited access to the values of physical/epistemic probabilities and causal probabilities. However, what makes these three kinds of probabilities useful is that when cognizers lack direct knowledge of them, they can still estimate them defeasibly using classical and nonclassical direct inference. These apply directly to physical/epistemic probabilities and nomic probabilities, but as I will show they also support defeasible inferences regarding causal probabilities.

It is useful to remember that the principles of classical and nonclassical direct inference are theorems of the theory of nomic probability — not primitive assumptions. They follow from the principle (SS) of the statistical syllogism and the calculus of nomic probabilities. Classical direct inference tells us how to infer the values of mixed physical/epistemic probabilities from the values of associated nomic probabilities. The core principle for computing the value of a conditional mixed physical/epistemic probability is:

(CDI*) If A is projectible with respect to B and C then "$\mathbf{W}(P \leftrightarrow Ac)$ & $\mathbf{W}(Q \leftrightarrow Cc)$ & $\mathbf{W}Bc$ & $\mathbf{prob}(Ax/Bx \,\&\, Cx) = r$" is a defeasible reason for "$\textsf{PROB}(P / Q) = r$".

In effect, this principle tells us that if $(P \leftrightarrow Ac)$ and $(Q \leftrightarrow Cc)$ are warranted, then we can identify

PROB(P / Q) with PROB(Ac / Cc), and if Bc is warranted we can defeasibly expect PROB(Ac / Cc) to be the same as **prob**$(Ax / Bx$ & $Cx)$. A consequence of this is that it is defeasibly reasonable to expect any further information we might acquire about c to be irrelevant to the value of PROB(Ac / Cc).

Nonclassical direct inference has a similar flavor, telling us that it is defeasibly reasonable to expect further projectible properties C to be statistically irrelevant to the nomic probability **prob**(Ax / Bx):

(DI) If A is projectible with respect to B then "**prob**$(Ax / Bx) = r$" is a defeasible reason for "**prob**$(Ax / Bx$ & $Cx) = r$".

From these two principles, we can derive a defeasible presumption of statistical irrelevance for mixed physical/epistemic probabilities:

(IR) For any P,Q,R, it is defeasibly reasonable to expect that PROB$(P / Q\&R)$ = PROB(P / Q).

This conclusion is forthcoming from the preceding principles of classical and non-classical direct inference. Given "$\mathbf{W}(P \leftrightarrow Ac)$ & $\mathbf{W}(Q \leftrightarrow Cc)$ & $\mathbf{W}(R \leftrightarrow Dc)$ & $\mathbf{W}Bc$" it is defeasibly reasonable to conclude that PROB(P / Q) = **prob**$(Ax / Bx$ & $Cx)$ and to conclude that PROB$(P / Q\&R)$ = **prob**$(Ax / Bx$ & Cx & $Dx)$. But by non-classical direct inference, it is also defeasibly reasonable to expect that **prob**$(Ax / Bx$ & Cx & $Dx)$ = **prob**$(Ax / Bx$ & $Cx)$, so it is defeasibly reasonable to expect that PROB(P / Q) = PROB$(P / Q\&R)$. Note that this immediately implies an analogous principle of irrelevance for causal probability:

(CIR) For any P,Q,R, it is defeasibly reasonable to expect that C-PROB$_A(P / Q\&R)$ = C-PROB$_A(P / Q)$.

Now let us apply these observations to the computation of causal probabilities. (IR) gives us a defeasible presumption that actions are not statistically relevant to their backgrounds, and by theorem 2 it follows that C-PROB$_A(O)$ = PROB(O / A). In other words, it is defeasibly reasonable to expect classical decision theory to yield the correct prescriptions. Causal decision theory only yields different prescriptions in the unusual case in which actions are statistically relevant to their backgrounds. These seem to be cases in which the action and the possible outcome have common causal ancestors. As has been noted repeatedly in the literature, these cases are unusual.

In a case in which an action is known to be statistically relevant to some elements of its backgrounds, it follows from (IR) that it is defeasibly reasonable to expect that all other elements of its backgrounds are statistically independent of the action in the strong sense required by theorem 2. Then it follows by theorem 2 that we can confine our attention to just those elements of the backgrounds that the action is known to be statistically relevant to, and hence deal with very small backgrounds. If more statistical relevance is found later, then the computation must be revised, but it is always defeasibly reasonable to expect that such a recomputation will not be necessary.

For some purposes, the preceding remarks put the cart before the horse. They suggest that $\textbf{C-PROB}_A(O)$ will be computed by first computing $\textbf{PROB}(O/A)$. In fact, I think the converse is likely to be true. The characterization of $\textbf{C-PROB}_A(O)$ in terms of scenarios provides what is in effect a recursive characterization, enabling us to compute causal probabilities by propogating them through time. If the case is classical, this is simultaneously a computation of $\textbf{PROB}(O/A)$. But if the case is not classical, i.e., actions affect the probabilities of events occuring earlier than themselves, that does not make the causal probabilities harder to compute, but it makes $\textbf{PROB}(O/A)$ much harder to compute.

14. Conclusions

Decision theory is an attempt to articulate principles governing the rational deliberations of cognizers faced with uncertainty. Most work in decision theory has been Bayesian, employing subjective probabilities. For the reasons given above, I regard subjective probability as a philosopher's fiction, inapplicable to the deliberations of real cognizers about what, rationally, they ought to do. However, decision theory can be reformulated using objective probabilities in place of subjective probabilities. Specifically, uncertainty can be represented using mixed physical/epistemic probabilities. Objective decision theory has a somewhat different flavor from Bayesian decision theory. In Bayesian decision theory the expected-values of actions directly dictate what the agent should do, but in objective decision theory it is *beliefs about* the expected-values that determine what the agent should do. This difference is not so marked when we move to causal decision theory, because theories like Skyrms' are formulated in terms of the objective concept of causal dependence. To make use of that, the agent must have beliefs about it. However, as Lewis observed, Skyrms' theory does not entirely accommodate this observation.

Some form of causal decision theory is required to handle the counterexamples to classical decision theory growing out of the Newcomb problem. As Lewis observes, the different causal decision theories that have been proposed are closely related to one another. In particular, they define causal probability by reference to concepts like causal dependence. Causal dependence is a philosophically problematic concept. I do not doubt that it makes sense, but its analysis is extremely controversial, as are its logical properties. Accordingly, it seems undesirable to use it as a primitive constituent of an analysis of causal probability.

This paper has proposed that within the context of objective causal decision theory we can replace the appeal to causal dependence by appeal to temporal relations and statistical relevance between mixed physical/epistemic probabilities. The basic idea is simply that causes propogate through the world in temporal order. The resulting analysis handles the known counterexamples to classical decision theory in essentially the same way Skyrms' theory does, but without appealing to vaguely understood concepts like causal dependence.

Bibliography

Gelfond, Michael, and Lifschitz, Vladimir

1993 "Representing action and change by logic programs", *Journal of Logic Programming* **17**, 301-322.

Gibbard, Alan and William Harper

1978 "Counterfactuals and two kinds of expected value", in *Foundations and Applications of Decision Theory*, ed. C. A. Hooker, J. J. Leach and E. F. McClennen, Reidel, Dordrecht, 125-162.

Hanks, Steve, and McDermott, Drew

1986 "Default reasoning, nonmonotonic logics, and the frame problem", AAAI-86.

1987 "Nonmonotonic logic and temporal projection", *Artificial Intelligence* **33**, 379-412.

Kyburg, Henry, Jr.

1974 *The Logical Foundations of Statistical Inference.* Dordrecht: Reidel.

1983 "The reference class". *Philosophy of Science* 50, 374-397.

Lewis, David

1981 "Causal decision theory", *Australasian Journal of Philosophy* **59**, 5-30.

Levi, Isaac

1977 "Direct inference". *Journal of Philosophy* **74**, 5-29.

1980 *The Enterprise of Knowledge.* Cambridge, Mass.: MIT Press.

1981 "Direct inference and confirmational conditionalization", *Philosophy of Science* **48**, 532-552.

Lifschitz, Vladimir

1987 "Formal theories of action", in M. L. Ginsberg (ed.), *Readings in Non-monotonic Reasoning*. Morgan-Kaufmann: Los Altos, CA.

Nozick, Robert

1969 "Newcomb's problem and two principles of choice", in *Essays in Honor of Carl Hempel*, ed.

N. Rescher, Reidel, Dordrecht, Holland, 107-133.

Pollock, John L.

1983 “A Theory of Direct Inference”, *Theory and Decision* **16**, 29-96.

1984 “Foundations for direct inference”, *Theory and Decision* **17**, 221-256.

1986 *Contemporary Theories of Knowledge*, Rowman and Littlefield.

1990 *Nomic Probability and the Foundations of Induction*, Oxford University Press.

1992 “The theory of nomic probability”, *Synthese* **90**, 263-300. An updated version of this paper is available at http://www.u.arizona.edu/~pollock.

1998 “Perceiving and reasoning about a changing world”, *Computational Intelligence.* **14**, 498-562.

Pollock, John, and Joseph Cruz

1999 *Contemporary Theories of Knowledge*, 2nd edition, Rowman and Littlefield.

Reichenback, Hans

1949 *A Theory of Probability*. Berkeley: University of California Press. (Original German edition 1935)

Shanahan, Murray

1990 “Representing continuous changes in the event calculus”. *ECAI-90*, 598-603.

1995 “A circumscriptive calculus of events”. *Artificial Intelligence*, **77**, 249-284.

1996 “Robotics and the common sense informatic situation”, *Proceedings of the 12th European Conference on Artificial Intelligence*, John Wiley & Sons.

1997 *Solving the Frame Problem*, MIT Press.

Shoham, Yoav

1986 *Time and Causation from the standpoint of artificial intelligence*, Computer Science Research Report No. 507, Yale University, New Haven, CT.

1987 *Reasoning about Change*, MIT Press.

Skyrms, Brian

1980 *Causal Necessity*, Yale University Press, New Haven.

1982 “Causal decision theory”, *Journal of Philosophy* **79**, 695-711.

1984 *Pragmatics and Empiricism*, Yale University Press, New Haven.

Sobel, Howard

1978 *Probability, Chance, and Choice: A Theory of Rational Agency*, unpublished paper presented at a workshop on Pragmatism and Conditionals at the University of Western Ontario, May, 1978.

Stalnaker, Robert

1978 “Letter to David Lewis”, reprinted in *Ifs*, ed. W. Harper, R. Stalnaker, G. Pearce, Reidel, Dordrecth, Holland.