# 6 Decision Making via Graphical Models

## 6.1 Roles for Causal Knowledge in Decision Making

The previous two chapters focused on using graphical models to represent our causal and conceptual knowledge. Those cognitive representations are, however, essentially impotent on their own; they are only useful if they are connected in some way with decision-making processes that can (intelligently) use them. The first section of this chapter aims to show that causal knowledge—represented as a DAG-based graphical model—plays a key role in much of our decision making. In particular, causal knowledge can guide us to attend to the proper factors and enable us to better predict the outcomes of our own actions. One type of decision making can thus be understood as operations on graphical models. Graphical models are not restricted, however, to serving as input to standard decision-making algorithms. In section 6.2, I describe two novel decision-making algorithms that are inspired and shaped partly by the advantageous features of their graphical model input. These novel algorithms still show only that graphical models can be an input to our decision making. I thus turn in section 6.3 to model the decision-making process itself as operations on decision networks (also called influence diagrams), a generalization of the DAG-based graphical models used in previous chapters. But first we look at the "easy" case: causal knowledge (structured as a DAG-based graphical model) provides a basis for intelligent decision making.

One important challenge in judgment and decision making is determining which factors should be the focus of one's attention; that is, my decision making requires that I pay attention to possibly relevant factors, rather than the definitely irrelevant ones. Many accounts of human decision making assume that this information search problem has already been solved,

as they assume that relevant features of the decision situation are provided as input. We can instead determine the relevant factors dynamically using causal knowledge, particularly if it is structured as a graphical model. In particular, suppose that my decision depends on some factor *T* that I do not know and thus need to predict. In that case, I should presumably focus on *T*'s direct causes and direct effects. This intuition has been explored extensively in the context of the "Take the Best" (TTB) heuristic for binary forced choices (Gigerenzer, 2000; Gigerenzer & Goldstein, 1996), and so we need a brief digression to explain TTB. It is important to recognize, however, that what matters is how causal knowledge can, in general, support information search in decision situations; the focus on TTB is a purely contingent matter. In particular, I will not discuss whether TTB is empirically correct, since we are principally interested in a functional role of causal knowledge (in information search) for decision contexts.

Consider making a choice between two options relative to some criterion, such as which washing machine is better or which city is larger. The basic idea of TTB is that people sequentially consider different cues—properties or features of the options—until they find a cue for which the options have different values (e.g., positive versus negative versus unknown), and then people choose the one with the "better" cue value. To take a simple example, suppose one is trying to decide which of two American cities is larger. One feature of American cities (that is related to size) is whether they have a professional (American) football team; TTB says (again, roughly) that if one city has a professional football team (positive cue value) but the other does not (negative cue value), then one should choose the city with a team as larger. Alternately, I might be trying to purchase a washing machine and decide between two options based solely on which is cheaper, or which has the longer warranty. Many papers provide empirical evidence that people behave roughly as predicted by TTB, though the precise empirical scope of TTB is a complicated matter (Broder, 2000; Gigerenzer, 2000; Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 1999; Newell & Shanks, 2003).

At the same time, it is somewhat puzzling that people would use TTB, as this strategy appears to be highly suboptimal. In TTB, decisions ultimately rest on only a single cue; it is, to use the jargon, noncompensatory: no amount of information in other cues can compensate for the single, critical difference. Thus, for example, one might judge Pittsburgh to be larger than Los Angeles, since Pittsburgh has a professional American football

team and Los Angeles does not, despite the numerous other cue differences telling us that Los Angeles is much larger. Perhaps surprisingly, however, the suboptimality of TTB is only apparent, not actual. In practice, TTB frequently produces close-to-optimal performance on various tasks (Chater, Oaksford, Nakisa, & Redington, 2003; Gigerenzer, Czerlinski, & Martignon, 1999; Hogarth & Karelaia, 2006; Martignon & Hoffrage, 1999), principally because cues are *not* considered in a random order. Rather, TTB assumes that people consider the different cues in order of their cue *validity*: the probability that a cue provides an accurate decision, given that it discriminates between the options at all. TTB works well precisely (and only) because the order in which people consider cues tracks their likelihood of giving the right answer (Newell & Shanks, 2003). This assumption that cues are considered in order of their validity is relatively innocuous in experimental settings, as participants are often told the cue validities (or are assumed to know them from life experience). In the real world, however, it is unclear how people identify possibly relevant cues, how they learn the appropriate validities, or whether they consider the cues in the proper order even if they have somehow learned the validities.

Causal knowledge potentially provides a solution to this problem of knowing what factors matter for a decision (Garcia-Retamero & Hoffrage, 2006; Garcia-Retamero, Wallin, & Dieckmann, 2007). In particular, causes and effects of the decision-relevant attribute $T$ are more likely to be valid cues for $T$ (Garcia-Retamero & Hoffrage, 2006). For example, having a professional American football team is an effect of a city's size, as a large population base is one cause of whether sufficient financial and political support exists for a team. Conversely, being a state capital is plausibly a cause of a city's size, as the presence of the state government causes there to be more jobs (all else being equal) and so more people. One might thus expect that both being a state capital and having a professional American football team would have significant, nonzero cue validities for city population, as they in fact do. On the other hand, the length of a city's name is not plausibly a cause or effect of the city's size, so we should expect that its validity would be essentially zero.

We thus seem to have a potential solution to the problem of the source of cue validity orderings and values: we use our causal knowledge to identify direct causes and effects and then compute cue validities for them (Garcia-Retamero & Hoffrage, 2006). We previously saw (in chap. 4) that our

causal knowledge seems to be structured as DAG-based graphical models. The identification problem is easily solved using such a graphical model, as the direct causes and effects are simply the factors that are (graphically) adjacent to *T*. And cue validities can easily be computed from the parametric information contained in the graphical model,[1] so we can order the cues correctly. Moreover, experiments have found that people perform substantially better on these types of binary forced-choice problems when they can use their causal knowledge, compared to having to estimate cue validities from observations (Garcia-Retamero et al., 2007); observed cue validity information is sometimes employed but requires some pretraining (Garcia-Retamero, Muller, Cantena, & Maldonado, 2009). It thus seems that TTB in real-world situations is plausibly an operation on a DAG-based graphical model. More generally, causal knowledge can provide the necessary structure for successful information search in decision situations.

We now turn to examining a role for causal knowledge in the actual decision-making process, not just as an attentional focusing mechanism. An almost-ubiquitous schema for cognitive models of decision making is that people (i) have a set of possible actions; (ii) consider for each action the various probabilistic outcomes of that action; and (iii) choose the action that has the best (probabilistic) value. There are, of course, many ways of instantiating this schema and many processing variants. However, this basic schema—evaluate the possible outcomes of different actions and choose the "best"—underlies almost all psychological models of choice (and most normative ones). One issue for any model of decision making is explaining step (ii): how does the decision maker determine the probabilistic outcomes of an action? One cannot simply compute (naive) probabilities conditional on the state resulting from the action, since those do not distinguish between causes and effects. For example, the probability of being over fifty conditional on having gray hair is *not* the same as the probability that I will be over fifty if I dye my hair gray. Actions on *T* presumably (though see sec. 6.3) only lead to probabilistic changes in *T*'s effects, not *T*'s causes, and so, to ensure correct predictions in step (ii), we need to know which associated factors are causes and which are effects (Joyce, 1999; Lewis, 1981).

The graphical-models-based approach can easily make sense of this critical distinction if I understand my decisions or actions as interventions that override (or "break") the normal causal structure. As we have seen several times earlier in the book, interventions on *T* break the edges into *T*, but not

the ones out of $T$ (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000), and this behavior exactly satisfies the requirement from the previous paragraph.[2] This complication did not arise when discussing TTB, since the decisions in that process are based on causal structure but do not change it. Both causes and effects can provide valuable cues for TTB, since both carry information. In contrast, actually acting in the world requires that we distinguish (correctly) between causes and effects in both learning and reasoning.

Thus, to the extent that people actually understand their choices as interventions, causal knowledge (represented as a DAG-based graphical model) can satisfy a critical role in decision making by providing the resources to solve step (ii). Substantial empirical evidence now shows that people do think about their own actions in this way. For example, if people believe that $C$ causes $E$, then they conclude that an action on $E$ will make no difference to $C$, while actions on $C$ will be a good means to bring about $E$ (Hagmayer & Sloman, 2009; Sloman & Hagmayer, 2006; Waldmann & Hagmayer, 2005). In addition, quantitative aspects of the causal knowledge matter: stronger causes are more likely to be chosen when people are trying to bring about some effect (Nichols & Danks, 2007). Of course, one might wonder whether the causal *knowledge* is really being used for decision making, rather than people relying on habits or implicit learning. It is surely the case that some "decisions" actually arise from habit and so might not be based on causal knowledge. Just as surely, not all decisions are habitual; many decisions are based on conscious deliberation, whether that reasoning occurs spontaneously (e.g., in novel situations) or in response to a prompt. Moreover, when people deliberate on their decision, then they make different, and better, choices than when simply responding immediately or automatically (Mangold & Hagmayer, 2011). This result suggests that many decisions are actually based on reasoning about the system. Finally, changes in one's beliefs about the causal structure lead to corresponding changes in decisions that cannot be explained simply on the basis of observed associations (Hagmayer & Meder, 2013). We thus have multiple threads of evidence that people's decision making frequently uses their distinctively causal knowledge to predict outcomes of interventions or actions. More importantly for me, this causal knowledge is best understood in terms of DAG-based graphical models; that is, decision making arguably depends (often) on exactly the shared representational store that underlies cognition about causation and concepts.

## 6.2 Novel Decision-Making Algorithms

The previous section focused on ways in which traditional, empirically supported decision-making algorithms can be understood as operations on cognitive representations of causal knowledge (structured as DAG-based graphical models). The graphical models framework can also suggest some novel ways to approach decision making, and I explore the computational aspects of two proposals here. Unfortunately, no empirical tests have occurred for either proposal, so we cannot determine whether either is descriptively correct. I thus focus more on the ways that novel decision-making theories can arise if cognitive representations are (approximately) graphical models.

The first novel approach starts with the observation that standard decision-making algorithms assume that people calculate, perhaps implicitly, the (probabilistic) outcomes of their actions to make a choice. In situations with complex causal structures, however, these computations can be quite complex, so we might hope to avoid them by using a suitable heuristic. Meder, Gerstenberg, Hagmayer, and Waldmann (2010) propose an Intervention-Finder heuristic: in trying to bring about some $E$, people consider each of the possible causes of $E$ (identified by statistical or other cues) and simply choose the option $C$ that maximizes the observed (not causal) conditional probability of $E$ given $C$, $P(E \mid C)$. That is, their heuristic uses causal knowledge to narrow the scope of deliberation to just the potential causes (rather than effects) but does not use that knowledge for the probability calculations. In particular, the Intervention-Finder heuristic does not attempt to account for common causes or other reasons why $P(E \mid C)$ might be large; it simply uses the observed conditional probability as a "good estimate" of what (probabilistically) would happen if one acted on $C$. Unsurprisingly, this heuristic works best when the different possible causes are not highly associated with one another; surprisingly, it performed close to optimally in a large-scale simulation study (Meder et al., 2010). Although the Intervention-Finder heuristic ignores causally important information, it turned out (in that simulation study) that the information being ignored did not usually make a difference as to which action was selected: the additional causal knowledge typically changed the value estimates of the actions, but not the overall rank order of them. This heuristic thus shows one way in which causal knowledge (structured as a DAG-based graphical

model) can suggest a novel decision-making algorithm that is normatively incorrect (since it does not actually try to determine the outcome probabilities) but nonetheless successful. Unfortunately no empirical studies have yet explored whether the Intervention-Finder heuristic is actually used by people, though the theoretical possibility is intriguing.

A somewhat more radical novel decision-making model emerges when we reflect on the assumption (of the standard decision-making schema) that the decision maker is presented with a set of actions to consider. More than fifty years ago, Duncan Luce (1959, pp. 3–4) wrote:

> There seems to have been an implicit assumption [in decision models] that no difficulty is encountered in deciding among what it is that an organism makes its choices. Actually, in practice, it is extremely difficult to know. … All of our choice theories—including this one—begin with the assumption that we have a mathematically well-defined set, the elements of which can be identified with the choice alternatives. How these sets come to be defined for organisms, how they may or may not change with experience, how to detect such changes, etc., are questions that have received but little illumination so far.

The situation arguably has not improved substantially since the publication of Luce's book. In response, one could flip the definition of a decision problem from being "action based" to "goal based": that is, a decision problem can potentially be characterized by a desired outcome state, rather than a set of possible actions. Goal-defined decision problems are surely not the only type that we face, but they seem to constitute a large set of everyday choices.

Given that we have only a desired outcome state and not a set of actions, we can use our causal knowledge—structured as a DAG-based graphical model—to dynamically construct possible choices. At a high level, the decision maker starts with the target variable $T$ and considers actions on its direct causes (according to her causal knowledge), given observations of the factors adjacent to $T$. That is, the decision maker first considers only those factors that her causal knowledge deems to be immediately relevant to $T$. If one of those actions is acceptable,[3] then the decision maker performs that action. If no actions are acceptable, then the decision maker moves her scope "out" one step on the basis of her causal knowledge: rather than considering actions that affect the direct causes (i.e., parents) of $T$, she considers actions that change the causes of the causes (i.e., grandparents) of $T$. And if none of *those* is acceptable, then she iterates back yet another step.

The process continues until the decision maker either finds an acceptable action or decides to stop for other reasons, whether boredom, frustration, or other demands on her cognitive resources. This method thus naturally constructs an appropriate set of possible actions, and so we seem to have the beginnings of a response to Luce's challenge.

Moreover, this algorithm can be extended in intriguing ways. Dynamic information gathering can be incorporated immediately, since, for any action $A$, the most relevant other factors that need to be considered are alternative causes of $A$'s descendants. That is, if I am considering an action on node $A$, I can easily use my causal graph to determine the other variables that might interact with my action to produce good or bad effects. A small adjustment to the algorithm thus yields a dynamic attentional focusing mechanism. Alternately, a challenge for decision making is recognizing and accounting for side effects of our choices. Such side effects can readily be determined for each action because the algorithm does not use a fixed variable set, though this additional computation does have some cost. Finally, planning behavior can be modeled by evaluating multiple actions rather than simply single ones, though that extension is still only hypothetical.

At any particular moment in the decision process, this method looks quite similar to the standard decision-making schema: the decision maker is considering the (probabilistic) impacts of a potential action to decide whether it is worth doing. Overall, however, there are several crucial differences. Perhaps most importantly, this decision method is fundamentally dynamic in scope, and so one's information-gathering and "action possibility" space are constantly changing. The decision maker never evaluates all possible actions but instead evaluates, and either accepts or rejects, each choice individually. The set of possible actions changes throughout the course of deliberation, as do the variables that are relevant for value judgments. Decision making thus shifts from the standard model of a process in which the decision maker is selecting from a menu of options to one in which the decision maker can dynamically explore and construct the very possibility set being evaluated. As a result, "choose the optimal action" cannot possibly be a decision criterion, since the decision maker will not necessarily ever know what constitutes "optimal." In many cases, the decision maker will be in a state of uncertainty about whether a better option potentially lurks over the horizon if she would consider just one more action. Roughly speaking, the agent can safely stop looking for better options only

if there is a "dominant" cause that is more efficacious than combinations of other changes and cannot reliably be brought about through indirect means.[4] There must inevitably be a satisficing aspect to many of the decisions made by this process: the eventual decision is selected because it is judged to be "good enough," not because it is known to be optimal (though it might be).

Despite these limitations, simulation tests that I conducted with Stephen Fancsali revealed that (a fully precise version of) this algorithm performs close to optimally using significantly fewer computations. For these simulations, we used a version of this algorithm that regards an action as "acceptable" if its value—operationalized as expected value—exceeds some threshold $\tau$; obviously, much more sophisticated versions are possible. We used two additional simplifications in these simulations: (i) when actions have costs, they depend only on the variable, not the value to which it is set; and (ii) the target $T$ being in the desired state is worth one unit of value, and all other results are worth zero. These assumptions mean that the value of an action is always less than or equal to one and can be negative (if the action is costly but unlikely to produce the desired state). We tested the algorithm on a large number of different causal structures under a wide range of parameter settings. I group together causal structures with the same number of variables (denoted by V), as this is a major driver of computational challenges. We otherwise simply average over performance for each graph size; the graphs in figures 6.1 through 6.3 provide mean performance for roughly 100,000 (V = 4), 300,000 (V = 6, 8, 10), or 10,000 (V = 15) graphs. Note that these results are for random graphs and parameterizations. There are obviously specific graphs on which this algorithm does quite poorly; I give an example later in this section when discussing empirical tests of this algorithm. One open question is what causal belief structures are "typical," and whether this algorithm performs particularly well on those graphs.

A first question is whether this algorithm typically finds the optimal action. Figure 6.1 plots the fraction of times that the algorithm selected the optimal action (vertical axis) against $\tau$, the value threshold for "acceptability" (horizontal axis). Unsurprisingly, the probability of choosing the optimal action increases as $\tau$ does; if one is choosier about the action, then one is more likely to wait until the truly optimal action is found. More interestingly, the mean match rate only exceeds 95 percent for $\tau = 0.8$. For small

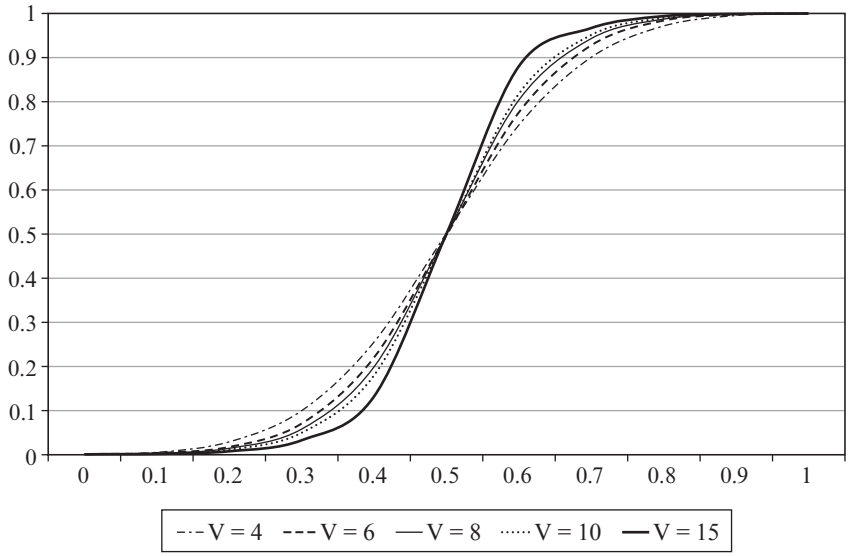


**Figure 6.1**
Match rate as a function of value threshold.

value thresholds—that is, if the algorithm is not particularly "choosy"—the algorithm rarely selects the optimal action.

An immediate follow-up question to this first analysis is: How good are the actions that *are* selected? Although the algorithm rarely selects the optimal action, is it choosing ones that are nonetheless reasonably good? Figure 6.2 shows the expected loss as a function of $\tau$; recall that the values of all choices are between zero and one (since their values must be less than or equal to one, and the algorithm can always choose to do nothing and receive at least zero units of value). Unsurprisingly, the mean loss decreases as $\tau$ increases: if one is more stringent about what one will accept, then one should ultimately make a better choice. More surprisingly, the algorithm performs quite well with only small losses in (absolute) expected value for small values of $\tau$. Even in the extreme case of $\tau = 0$ (i.e., accept the first considered action with a positive expected value), the algorithm performed reasonably well. At the upper end, the mean expected value loss for $\tau = 0.6$ was less than 0.01. This latter result is particularly notable given that the algorithm only chooses the optimal action 80 percent of the time when $\tau = 0.6$.

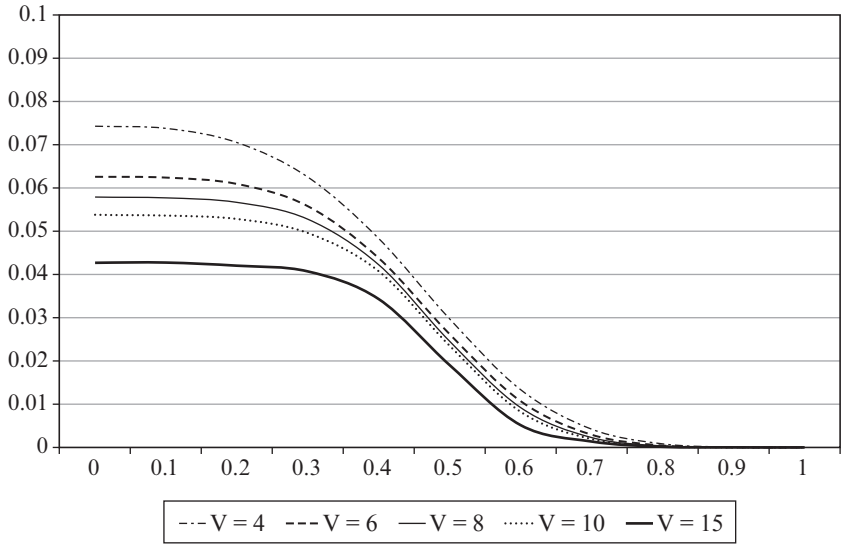


**Figure 6.2**
Expected loss as a function of value threshold.

Part of the cognitive plausibility of this algorithm is that it appears to be computationally much simpler. To confirm this idea, we compared the number of actions checked by the algorithm (including the "null" action of doing nothing) and the number of actions that were checked in finding the optimal choice. Figure 6.3 shows the former divided by the latter as a function of $\tau$, so smaller numbers indicate that the present algorithm considered fewer possible actions. Note that this measure ignores memory and other costs involved in information search when trying to find the optimal choice. A value of 1.0 indicates that optimal search and this algorithm evaluated the same number of actions; in these simulations, that point generally occurred around $\tau = 0.6$. One might be surprised to see numbers greater than 1.0. These are possible because the present algorithm collects information only gradually and thus rechecks the null action ("do nothing") after each new observation, while the optimal search algorithm checks it only once.

The key to these simulations is to consider figures 6.1 through 6.3 together, as they show the following two points: for small values of $\tau$, this algorithm is computationally much simpler (10–30 percent of the computational cost) and achieves close-to-optimal performance; for large values

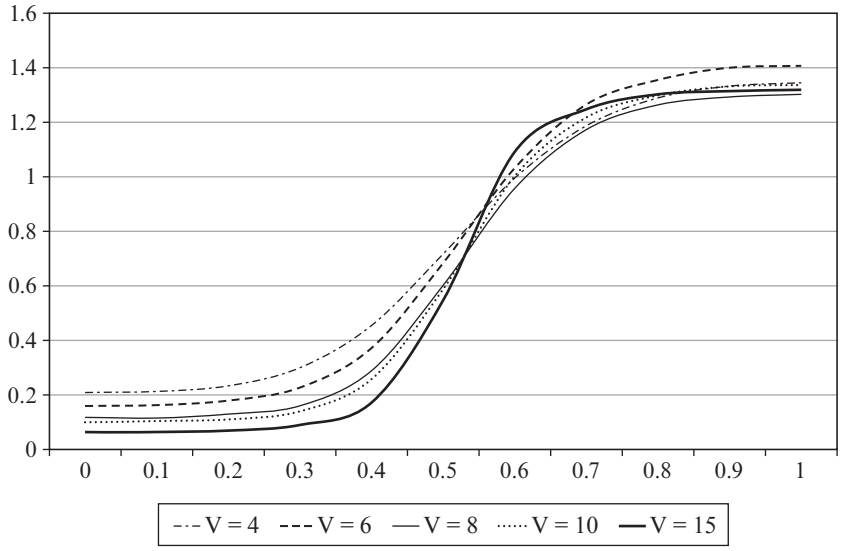


**Figure 6.3**
Fraction of actions checked relative to optimal algorithm.

of $\tau$, the algorithm is only slightly more complicated and successfully finds the optimal choice. Perhaps most importantly, as V gets larger, the computational advantage increases, and the expected value loss decreases. Given that people's causal knowledge presumably ranges over thousands of variables (or more), it is important that the performance trends of this algorithm are headed in the correct direction. Moreover, these changes as V increases are unsurprising, as this algorithm is explicitly designed to first check the options that are most likely to be best, while the optimal algorithm must check every conceivable option. By exploiting the relevance relations encoded in the causal knowledge graphical model, this algorithm exhibits a performance profile that fits the decision maker's needs and can usefully and sensibly be tuned as appropriate.

Of course, these simulations do not tell us that people's decision making actually uses something like this algorithm. I thus close this section by considering three ways to empirically test the algorithm. The real keys here, though, are not the particular experiments but (a) the way that the graphical models framework inspires a novel decision-making algorithm, and (b) that these experiments would arguably not be conducted without thinking

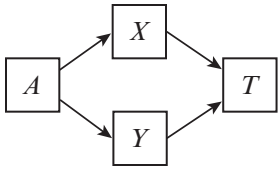


**Figure 6.4**
Structure to test decision-making algorithm.

about decision making through a graphical model lens. The first experiment looks at people's information acquisition patterns when confronted with a novel choice problem. Specifically, this proposed algorithm predicts that people will seek out information about direct causes before indirect causes. This prediction can be tested straightforwardly by teaching people a novel causal structure and then introducing a cost to information retrieval (e.g., forcing people to move a mouse over an on-screen box to learn a variable value). We can then investigate the patterns of information search that people exhibit.

A second experiment focuses on the dynamic "satisficing" aspect of the algorithm. More specifically, suppose that people learn, either from experience or from description, the causal structure in figure 6.4. This graphical model can have natural causal strength parameters that imply that $A$ is a more efficacious action than either $X$ or $Y$ alone. For example, suppose $X$ and $Y$ are both relatively weak causes of $T$, but $A$ is a strong cause of both of them. In this case, acting on $A$ makes $T$ more likely than an action on $X$ or $Y$ alone, precisely because there are two causal pathways from $A$ to $T$. The present algorithm predicts, however, that people should sometimes act suboptimally in cases such as these. That is, there should be conditions in which people will choose to do either $X$ or $Y$ rather than $A$, such as when there is time pressure or when the stakes are very low (and so people might be willing to perform actions even when they are not obviously close to optimal).

A third experiment is similar but instead uses the causal structure in figure 6.5, where we additionally assume that $X$ is a deterministic effect of $A$—$X$ occurs if and only if $A$ does—and that $Y$ is relatively ineffective. Because $X$ is a deterministic effect of $A$, actions on $A$ and $X$ result in exactly the same probability of $T$. If actions on $A$ and $X$ have the same cost, then almost all models of decision making predict that people should choose

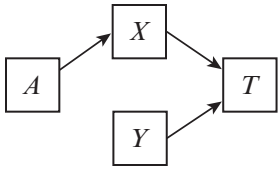


**Figure 6.5**
Alternative structure to test decision-making algorithm.

either randomly between $A$ and $X$ or on the basis of value-irrelevant factors (e.g., perceptual salience). If actions on $A$ and $X$ have identical impacts, then we seem to have no basis for picking one rather than the other. At the least, it is straightforward to develop experimental stimuli for which "always act on $X$" is not the predicted action. In contrast, the present algorithm predicts that people will always choose to intervene on $X$. Of course, for all these experiments, it is important that people learn the causal structure without reflecting deeply on it; if they have ample time to think about all the possibilities, then even the present algorithm predicts that people should make optimal choices. Nonetheless we have straightforward empirical tests for a novel decision-making algorithm prompted by modeling our causal knowledge as a DAG-based graphical model.

## 6.3 Graphical-Model-Based Decision Making

The previous two sections showed ways in which decision making can use graphical-model-based cognitive representations, both in the context of traditional models of decision making and to suggest novel algorithms (that remain to be empirically tested). Those demonstrations do fall short, however, of what was shown in chapters 4 and 5, where graphical models played an integral role in understanding the very cognitive processes; those representations were not simply (optional) inputs to the process. This section aims to show that graphical models can similarly play a central role in decision making. The DAG-based graphical models in the previous sections treated every node identically: they all represented variables that could take different values. The factors that are relevant to a decision-making situation are not, however, all of the same type. Rather, it seems natural to distinguish four different types of things that we would like to represent in our model: actions, values, factors in the world, and the decision-making

method itself. The variables that we have seen all along in our graphical models correspond to the factors in the world; to incorporate the other three, we need to expand the representational language of our graphical models.

Decision networks, also called influence diagrams, are DAG-based graphical models containing four distinct node types corresponding to the four types listed earlier (Howard & Matheson, 1981; Shachter, 1986). I will follow standard notational conventions and represent the different node types using different shapes. Nodes corresponding to events or factors in the world (i.e., outside the decision agent herself) are drawn in ovals, and all edges into them correspond to causal connections. Nodes representing the actions taken by the decision maker are drawn in rectangles; edges into such nodes capture informational or cognitive connections. Value nodes are denoted by diamond shapes, and edges into them indicate the factors that determine the value that the decision maker receives at a moment in time. Finally, the decision-making process itself is drawn in a triangle and is typically required not to have edges into it. This is all quite complex in the abstract, so it is perhaps easiest to understand by working through an example. Consider the simple problem of deciding whether to turn on the lights in a room with windows, where we assume that the decision maker wants to be able to see but does not want to waste electricity. A plausible decision network for this situation is shown in figure 6.6.

*Time of day* encodes whether it is daytime or nighttime, which is obviously outside the control of the decision maker. *Max utility* indicates that the decision maker is using a decision process in which she attempts to maximize her utility; for probabilistic decisions, one could instead use a strategy such as "choose the action that has the maximal probability of yielding a utility of at least *U*." The *Flip switch* node represents the action

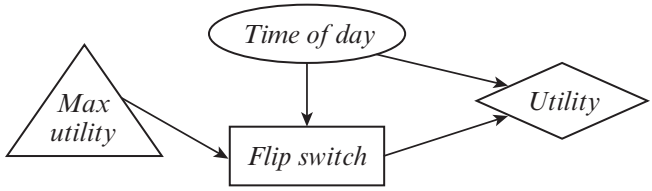


**Figure 6.6**
Example decision network.

that the decision maker actually takes, either "yes" or "no." That action is determined by the decision-making process (*Max utility*) given information from *Time of day*. Finally, *Utility* is a function of the time of day and the action, where her value presumably depends on (a) being able to see, but with (b) a cost for using electricity.

If we have a fully specified decision network, then we can use information about the world to derive a probability distribution for the possible actions (perhaps concentrated entirely on one choice, if we use a decision rule such as "maximize expected utility"). More generally, a decision network provides a representation of the decision situation that is faced by the decision maker, and so it encodes (in a graphical model) exactly the information required to make a decision in accordance with the specified decision rule (see also Solway & Botvinick, 2012). By providing additional information, a decision network goes beyond the types of graphical models that we have seen in previous chapters. For example, consider trying to represent the decision problem of "should I turn on the lights?" in an "ordinary" DAG-based graphical model. The causal structure is clear: *Light switch* $\rightarrow$ *Ability to see* $\leftarrow$ *Time of day*. But that causal graph tells us nothing about what I value, my abilities or capacities for action, or how I make decisions. These additional node types are required to make sense of decisions using just a graphical model.

These observations might seem to be in tension with sections 6.1 and 6.2, since I argued there that decision making could be understood using "ordinary" DAG-based graphical models. In those sections, however, the graphical model was actually *not* sufficient to explain the full decision-making process; it only captured one informational component or cognitive representation, rather than all elements of the process. The other relevant pieces (e.g., value function, decision-making process, etc.) were contained in processes or algorithms that sat outside the DAG-based graphical model, though they obviously used that graphical model. In contrast, the decision network framework enables us to put all those pieces into a single graphical model, thereby providing a much stronger graphical-model-based unification of the relevant decision-making representations. For example, the iterative decision search algorithm discussed in section 6.2 operated on causal knowledge but also used values, costs, thresholds, and other elements that were not in the causal graphical model. Using decision networks, we can represent almost all the pieces of that algorithm in a single graphical

model: (a) add a single value node that is the child of only the target variable; (b) for every variable on which I can act, include an action node that is its parent; and (c) add a single decision rule node that is the parent of every action node and has the iterative procedure as its value. The resulting decision network makes the same predictions as the algorithm from section 6.2, though at the cost of including more graphical nodes. These additions are worthwhile, however, precisely because they force us to be clear about the relevant elements, and also reveal assumptions (e.g., only single actions are allowed) that could naturally be relaxed. Similar characterizations can be provided for all of sections 6.1 and 6.2, not just the iterative decision algorithm.

All these observations about decision networks leave unaddressed the key descriptive question: do people's actual decisions suggest that they represent particular decision situations using decision networks? This question turns out to be trickier to answer than we might have expected, as most standard decision-making experiments do not test predictions that are relevant to this question. Instead those experiments typically focus on how people determine the value of a choice with multiple attributes, or how those values are used in choice, or both. Put into the language of decision networks, those experiments aim to determine the settings of the decision process and value nodes, rather than understanding and exploring the broader causal/informational structure in which they reside. At the same time, experiments that estimate those values are easily captured using decision networks precisely because they do not specify the settings of those nodes a priori. Standard decision-making experiments thus help to estimate some "free parameters" in the decision network graphical model, rather than testing the overall viability of the framework. To see whether people's choices can be understood as if they represent situations using decision networks, we need to find choice problems with more complex causal/informational structure.

Because the decision networks framework is still quite new in cognitive science, relatively few experiments have directly investigated it. Some such experiments do exist that explore distinctive predictions of decision networks, however, and they suggest that decision networks provide good (graphical) models of our cognitive representations of some decision situations. One set of experiments explores the ways in which decision networks can support inferences about the preferences of *other* people, given

observations of their choices. That is, if people understand their own choice situations using decision networks, then they plausibly represent others' choice situations in the same way. And just as decision networks can be used to generate my own choices, they can also provide the foundation for "backward" inference about others, from their choices to their preferences. For example, Jern and Kemp (2011) looked at inferences about the complex, structured preferences of others from pairwise choices over sets of objects. They found that participants' inferences about others' preferences (based on the choices that the others made) corresponded closely to the inferences that result from trying to estimate parameters in a decision network. That is, the conclusions we draw about what someone else likes or dislikes can naturally be understood as inferences about a decision network that models the other person's choices.

A second, and more interesting, set of experiments examines how we use observations of someone's choices to learn more about what they know and can do (Jern, Lucas, & Kemp, 2011). For example, suppose someone is playing a game in which the computer displays five letters on a screen, and the player has to write down a sixth letter that is different from the others. Additionally, suppose that a cover over part of the screen can obscure one or more of the letters. If I observe only the six letters (five from the computer, one from the contestant), can I learn the shape of the cover? Intuitively, the answer is clearly yes: the contestant would never knowingly choose a letter that matches one of the computer's letters (assuming she understands the game, responds sensibly, and so forth), and so if there is such a match, then that letter must be covered on the screen. Standard models of decision making cannot easily handle this type of situation, since they do not have a transparent representation of these types of informational links. In contrast, the player's choice problem can readily be modeled as a decision network, where the structure of the cover implies missing informational links. In fact, we can represent all the different decision networks that correspond to all the different covers that could be placed over the screen. Learning the structure is then just a matter of learning which decision network is the actual one; that is, learning is determining which decision network best predicts the player's actual choices. Jern et al. (2011) performed experiments structurally similar to this one and found that people gave responses that closely tracked the decision network predictions. Although experiments about decision networks in cognition have only just begun,

these results—coupled with the ability of decision networks to capture many standard results—give reason to suspect that many of our decisions are made using cognitive representations much like decision networks.

A general concern, however, lingers about *any* approach (including decision networks) that attempts to model choices as occurring within the causal structure of the world. From the decision maker's perspective, choice is almost always understood as exogenous and lying outside the relevant causal structure; if I think that my decisions are interventions (which has been repeatedly demonstrated), then I cannot think that they are caused by factors in the decision situation. Moreover, most models of decision making mirror this stance: standard decision theory, for example, represents the decision maker's choice as "free" and unconstrained by the other represented factors. From an outside observer's perspective, however, one can frequently understand a choice as being caused by other factors in the environment. For example, the outside temperature is clearly a (partial) cause of whether I decide to wear a sweater to work. In fact, whole research programs center on understanding the ways in which external factors (rather than rational deliberation) can cause our choices (e.g., the work described in Ariely, 2008). Whether anything causes choice thus depends on one's perspective (Sloman & Hagmayer, 2006), and uncertainty about the proper perspective might underlie self-deception phenomena (Fernbach, Hagmayer, & Sloman, 2014; Sloman, Fernbach, & Hagmayer, 2010). Moreover, this ambiguity and uncertainty raise concerns about the coherence between deliberation/choice and self-prediction (Levi, 1997, 2000, 2007; Rabinowicz, 2002), and even of standard decision theory itself (Kusser & Spohn, 1992), though decision theorists have resources with which to respond (Joyce, 2002). More generally, the worry is that graphical models, and decision networks in particular, cannot possibly provide a model of our cognitive representations about choice if we have no coherent understanding of choice in the first place.

The fundamental incompatibility between these two perspectives on the causal status of choice—outside versus inside the causal structure of the decision situation—arguably underlies the difficulties posed by Newcomb-type problems (Glymour & Meek, 1994; Hagmayer & Sloman, 2006). Newcomb's problem (Nozick, 1970) supposes that a decision maker is confronted with a seemingly simple choice between receiving the contents of (i) one covered box or (ii) that same covered box plus a transparent box

that has $1,000 in it. The tricky part of the problem comes in the way that the contents of the covered box are determined: an incredibly accurate (in the past) predictor places $1,000,000 in that box if she predicts that the decision maker will make choice (i), but nothing at all if the prediction is choice (ii). If we think about choice as exogenous, then it seems clear that the decision maker should choose both boxes. The prediction has already been made at the time of choice, and so the money is either in the covered box or not. Since the choice is outside the causal structure of the world, the *actual* decision cannot change whether the money is there. One should thus take both boxes, thereby receiving $1,000 more than whatever is in the covered box. If we instead adopt the perspective of choice as part of the causal structure of the world, then we understand our choices as presumably caused by our earlier selves. Those earlier selves can be known by the predictor, and so one's actual decision is *not* independent of whether the money is there; rather, both are effects of a common cause—namely, one's earlier self. Thus, given that the predictor has been accurate in the past, the decision maker should make choice (i) to maximize the chance that $1,000,000 is in the covered box.

Our responses to Newcomb's problem depend on whether we understand choice as exogenous or endogenous—outside or inside the causal structure of the world, respectively. These perspectives are incompatible with each other, however: we cannot simultaneously understand our choices as *both* exogenous *and* endogenous but rather must choose a perspective at a moment in time. As a result, Newcomb's problem has been thought to pose a challenge to all precise models of decision, whether based on graphical models or not. I suggest, however, that graphical models are particularly well situated to respond to Newcomb's problem, precisely because the perspective that one takes in a moment is transparently represented in the corresponding graphical model. In a decision network, actions are always understood to be "regular" causal nodes that can be influenced or shaped by other factors (e.g., the temperature outside). If choice is understood as endogenous, then there is nothing more to be done; the representation already captures the idea that my actions are caused by the world around me. If, however, choice has an exogenous component, then we simply add a decision process node with an edge into the action node; the addition of such a node exactly encodes the idea that some part of choice lies outside the rest of the causal structure (since decision process nodes never have

edges into them). These two perspectives on choices are truly irreconcilable, so no single framework can represent both simultaneously. The best that we can ask is that the framework clearly represent each possibility, as well as the different inferences that can be drawn in each. The graphical models framework does exactly that (see also Fernbach et al., 2014).

Over the course of the past three chapters, we have seen how the graphical models framework can represent the knowledge structures that underlie three major areas of cognition: causal learning and reasoning, concept acquisition and application, and decision making of different types. The next chapter turns to the challenge of putting these pieces together into a single cognitive architecture.