Chapter 2

Normative models of judgment and decision making

Jonathan Baron, University of Pennsylvania[1]

Pre-publication version of:

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*, pp. 19–36. London: Blackwell.

2.1 Introduction: Normative, descriptive, and prescriptive

The study of judgment and decision making (JDM) is traditionally concerned with the comparison of judgments to standards, standards that allow evaluation of the judgments as better or worse. I use the term "judgments" to include decisions, which are judgments about what to do. The major standards come from probability theory, utility theory, and statistics. These are mathematical theories or "models" that allow us to evaluate a judgment. They are called normative because they

[1] baron@psych.upenn.edu

are norms.[2]

This chapter is an introduction to the main normative models, not including statistics. I shall try to develop them informally, taking a more philosophical and less mathematical approach than other writers. Anyone who wants a full understanding should work through the math, for which I provide citations.

One task of our field is to compare judgments to normative models. We look for systematic deviations from the models. These are called biases. If no biases are found, we may try to explain why not. If biases are found, we try to understand and explain them by making descriptive models or theories. With normative and descriptive models in hand, we can try to find ways to correct the biases, that is, to improve judgments according to the normative standards. The prescriptions for such correction are called prescriptive models. Whether we say that the biases are "irrational" is of no consequence. If we can help people make better judgments, that is a good thing, whatever we call the judgments they make without our help.

Of course, "better" implies that the normative models truly define what better means. The more certain we are of this, the more confidence we can have that our help is really help. The history of psychology is full of misguided attempts to help people, and they continue to this day. Perhaps our field will largely avoid such errors by being very careful about what "better" means. If we can help people, then the failure to do so is a harm. Attention to normative models can help us avoid the errors of omission as well as those of commission.

In sum, normative models must be understood in terms of their role in looking for biases, understanding these biases in terms of descriptive models, and developing prescriptive models (Baron, 1985).

As an example, consider the sunk cost effect (Arkes & Blumer, 1985). People throw good money after bad. If they have made a down payment of $100 on some object that costs and additional $100, and they find something they like better for $90 total, they will end up spending more for the

[2]The term "normative" is used similarly in philosophy, but differently in sociology and anthropology, where it means something more like "according to cultural standards."

object they like less, in order to avoid "wasting" the sunk cost of $100. This is a bias away from a very simple normative rule, which is, "Do whatever yields the best consequences in the future." A prescriptive model may consist of nothing more than some instruction about such a rule. (Larrick et al., 1990, found such instruction effective.)

In general, good descriptive models help create good prescriptive models. We need to know the nature of the problem before we try to correct it. Thus, for example, it helps us to know that the sunk-cost effect is largely the result of an over-application of a rule about avoiding waste (Arkes, 1996). That helps because we can explain to people that this is a good rule, but is not relevant because the waste has already happened.

The application of the normative model to the case at hand may be challenged. A critic may look for some advantage of honoring sunk costs, which might outweigh the obvious disadvantage, within the context of the normative model. In other cases, the normative model is challenged. The fact that theories and claims are challenged does not imply that they are impossible to make. In the long run, just as scientific theories become more believable after they are corrected and improved in response to challenges, so, too, may normative models be strengthened. Although the normative models discussed in this chapter are hotly debated, others, such as Aristotle's logic, are apparently stable (if not all that useful), having been refined over centuries.

2.1.1 The role of academic disciplines

Different academic disciplines are involved in the three types of models. Descriptive models are clearly the task of psychology. The normative model must be kept in mind, because the phenomenon of interest is the deviation from it. This is similar to the way psychology proceeds in several other areas, such as abnormal psychology or sensation and perception (where, especially recently, advances have been made by comparing humans to ideal observers according to some model).

Descriptive models account not only for actual behavior but also for reflective judgments. It is possible that our reflective intuitions are also biased. Some people, for example, may think that it is correct to honor sunk costs. We must allow the possibility that they are, in some sense, incorrect.

The prescriptive part is an applied field, like clinical psychology, which tries to design and test ways of curing psychological disorders. (The study of perception, although it makes use of normative models, has little prescriptive work.) In JDM, there is no single discipline for prescriptive models. Perhaps the closest is the study of decision analysis, which is the use of decision aids, often in the form of formulas or computer programs, to help people make decisions. But education also has a role to play, including simply the education that results from "giving away" our findings to students of all ages.

Normative models are properly the task of philosophy. They are the result of reflection and analysis. They cannot depend on data about what people do in particular cases, or on intuitions about what people ought to do, which must also be subject to criticism.. The project of the branch of JDM that is concerned with normative models and biases is ultimately to improve human judgment by finding what is wrong with it and then finding ways to improve it. If the normative models were derived from descriptions of what most people do or think, we would be unable to find widespread biases and repair them.

Although the relevant philosophical analysis cannot involve data about the judgment tasks themselves, it must include a deeper sort of data, often used in philosophy, about what sort of creatures we are. For example, we are clearly beings who have something like beliefs and desires, and who make decisions on the basis of these (Irwin, 1971). A normative model for people is thus unlikely to serve as well for mosquitoes or bacteria.

2.1.2 Justification of normative models

How then can normative models be justified? I have argued that they arise through the imposition of an analytic scheme (Baron, 1994, 1996, 2000). The scheme is designed to fit the basic facts about who we are, but not necessarily to fit our intuitions.

Arithmetic provides an example (as discussed by Popper, 1962, who makes a slightly different point). The claim that $1 + 1 = 2$ is a result of imposing an analytic frame on the world. It doesn't seem to work when we add two drops of water by putting one on top of the other. We get one big

drop, not two. Yet, we do not say that arithmetic has been dis-confirmed. Rather, we say that this example does not fit our framework. This isn't what we mean by adding. We maintain the simple structure of arithmetic by carefully defining when it applies, and how.

Once we accept the framework, we reason from it through logic (itself the result of imposition of a framework). So no claim to absolute truth is involved in this approach to normative models. It is a truth relative to assumptions. But the assumptions, I shall argue, are very close to those that we are almost compelled to make because of who we are. In particular, we are creatures who make decisions based on beliefs and (roughly) desires.

Acts, states, and consequences

One normative model of interest here is expected-utility theory (EUT), which derives from an analysis of decisions into acts, uncertain states of the world, and consequences (outcomes). We have beliefs about the states, and desires (or values, or utilities) concerning the consequences. We can diagram the situation in the following sort of table:

	State X	State Y	State Z
Option A	Outcome 1	Outcome 2	Outcome 3
Option B	Outcome 4	Outcome 5	Outcome 6

The decision could be which of two trips to take, and the states could be the various possibilities for what the weather will be, for example. The outcomes could describe the entire experiences of each trip in each weather state. We would have values or utilities for these outcomes. EUT, as a normative model, tells us that we should have probabilities for the states, and that the expected utility of each option is determined from the probabilities of the states and the utilities of the outcomes in each row.

Before I get into the details, let me point out that the distinction between options and states is the result of a certain world view. This view makes a sharp distinction between events that we control (options) and events that we do not control (states). This view has not always been accepted.

Indeed, traditional Buddhist thought tries to break down the distinction between controllable and uncontrollable events, as does philosophical determinism. But it seems that these views have had an uphill battle because the distinction in question is such a natural one. It is consistent with our nature.

Another important point is that the description of the outcomes must include just what we value. It should not include aspects of the context that do not reflect our true values, such as whether we think about an outcome as a gain or a loss (unless this *is* something we value). The point of the model is provide a true standard, not to find a way to justify any particular set of decisions.

Reflective equilibrium

An alternative way of justifying normative models is based on the idea of "reflective equilibrium" (Rawls, 1971). The idea comes most directly from Chomsky (1957; see Rawls, 1971, p. 47), who developed his theory of syntax on the basis of intuitions about what was and what was not a sentence of the language. Rawls argues that, like the linguists who follow Chomsky, we should develop normative theories of morality (a type of decision making) by starting with our moral intuitions, trying to develop a theory to account for them, modifying the theory when it conflicts with strong intuitions, and ultimately rejecting intuitions that conflict with a well-supported theory.

Such an approach makes sense in the study of language. In Chomsky's view, the rules of language are shaped by human psychology. They evolved in order to fit our psychology abilities and dispositions, which, in turn, evolved to deal with language.

Does this approach make sense in JDM? Perhaps as an approach to descriptive theory, yes. This is, in fact, its role in linguistics as proposed by Chomsky. It could come up with a systematic theory of our intuitions about what we ought to do, and our intuitions about the judgments we ought to make. But our intuitions, however systematic, may be incorrect in some other sense. Hence, such an approach could leave us with a normative model that does not allow us to criticize and improve our intuitions.

What criterion could we use to decide on normative models? What could make a model incorrect? I will take the approach here (as does Over in ch. 1) that decisions are designed to achieve goals, to bring about outcomes that are good according to values that we have. And other judgments, such as those of probability, are subservient to decisions. This is, of course, an analytic approach. Whatever we call what it yields, it seems to me to lead to worthwhile questions.

2.2 Utility (good)

The normative models of decision making that I shall discuss all share a simple idea: the best option is the one that does the most good. The idea is that good, or goodness, is "stuff" that can be measured and compared. Scholars have various concepts of what this stuff includes, and we do not need to settle the issue here. I find it useful to take the view that good is *the extent to which we achieve our goals* (Baron, 1996). Goal achievement, in this sense, is usually a matter of degree: goals can be achieved to different extents. Goals are *criteria* by which we evaluate states of affairs, more analogous to the scoring criteria used by judges of figure-skating competitions than to the hoop in a basketball game. The question of "what does the most good" then becomes the question of "what achieves our goals best, on the whole."

If this question is to have meaningful answers, we must assume that utility, or goodness, is *transitive* and *connected*. Transitivity means that if A is better than B (achieves our goals better than B, has more utility than B) and B is better than C, then A is better than C. This is what we mean by "better" and is, arguably, a consequence of analyzing decisions in this way. Connectedness means that, for any A and B, it is always true that either A is better than B, B is better than A, or A and B are equally good. There is no such thing as "no answer." In sum, connectedness and transitivity are consequences of the idea that expected utility measures the extent to which an option achieves our goals. Any two options either achieve our goals to the same extent, or else one option achieves our goals better than the other; and if A achieves our goals better than B, and B

achieves them better than C, then it must be true that A achieves them better than C.[3]

Sometimes we can judge directly the relation between A and B. In most cases, though, we must deal with trade-offs. Option A does more good than B in one respect, and less good in some other respect. To decide on the best option, we must be able to compare *differences* in good, i.e., the "more good" with the "less good." Mathematically, this means that we must be able to measure good on an interval scale, a scale on which intervals can be ordered.

Connectedness thus applies even if each outcome (A and B) can be analyzed into parts that differ in utility. The parts could be events that happen in different states of the world, happen to different people, or happen at different times. The parts could also be attributes of a single outcome, such as the price and quality of a consumer good.

Some critics have argued that this is impossible, that some parts cannot be traded off with other parts to arrive at a utility for the whole. For example, how do we compare two safety policies that differ in cost and number of deaths prevented? Surely it is true descriptively that people have difficulty with such evaluations. The question is whether it is reasonable to assume, normatively, that outcomes, or "goods" can be evaluated as wholes, even when their parts provide conflicting information.

One argument that we can assume this is that sometimes the trade-offs are easy. It is surely worthwhile to spend $1 to save a life. It is surely not worthwhile to spend the gross domestic product of the United States to reduce one person's risk of death by one in a million this year. In between, judgments are difficult, but this is a property of all judgments. It is a matter of degree. Normative models are an idealization. The science of psychophysical scaling is built on such judgments as,

---

[3]Another way to understand the value of transitivity is to think about what happens if you have *in*transitive preferences. Suppose X, Y, and Z are three objects, and you prefer owning X to owning Y, Y to Z, and Z to X. Each preference is strong enough so that you would pay a little money, at least 1 cent, to indulge it. If you start with Z (that is, you own Z), I could sell you Y for 1 cent plus Z. (That is, you pay me 1 cent, then I give you Y, and you give me Z.) Then I could sell you X for 1 cent plus Y; but then, because you prefer Z to X, I could sell you Z for 1 cent plus X. If your preferences stay the same, we could do this forever, and you will have become a *money pump*.

"Which is larger, the difference between the loudness of tones A and B or the difference between B and C?" When subjects in experiments make a large number of such judgments, their average responses are orderly, even though any given judgment feels like a wild guess. This is, arguably, the sort of creatures we are. We have some underlying order, wrapped in layers of random error. (See Broome, 1997, and Baron, 2002, for related arguments.)

Another sort of challenge to the idea of a single utility for whole outcomes is that utility judgments are easily influenced by extraneous manipulations. I have argued that all of these manipulations do not challenge utility as a normative ideal (Baron, 2002). The general argument is that it is possible to understand the effects of manipulations as distortions of a true judgment.

On the other hand, utilities change as a result of reflection. They are not hard wired, and the theory does not require them to be. They are best seen as something more like concepts, formed on the basis of reflection, and constantly being modified (Baron, 2002).

2.3 Expected-utility theory (EUT)

Expected-utility theory (EUT) deals with decisions under uncertainty, cases in which we analyze outcomes into parts that correspond to outcomes in different states of the world. The theory says that the overall utility of an option is the expected utility. That is, the utility averaged across the various possible states, with the outcomes weighted according to the probability of the states. It is analogous to calculating the average, or expected, winning from a gamble. If you get \$12 when a die comes up with a 1 and \$0 otherwise, the average winning is \$2, because the probability of a 1 is 1/6. But EUT deals with utility, not money. The mathematical and philosophical basis of this theory developed in the 20th century (Ramsey, 1931; de Finetti, 1937; von Neumann & Morgenstern, 1947; Savage, 1954; Krantz et al., 1970; Wakker, 1989).

Table 2.1 shows an example of several bets, with A and B being the uncertain states. These could be whether a coin is heads or tails, or whether it rains tomorrow or not. The outcomes in each cell are gains in dollars. The expected utility (EU) of an option is computed, according to

EUT, by multiplying the utility (U) of each outcome by its probability (p), and then summing across the possible outcomes. We would thus have to assign a probability to states A and B. The EU of option S is thus $p(A)U(300) + p(B)U(100)$. To decide between options S and T we would ask which has greater EU, so we would look at the difference. This amounts to

$$[p(A)U(\$300) + p(B)U(\$100)] - [p(A)U(\$420) + p(B)U(\$0)]$$

or

$$p(A)[U(\$300) - U(\$420)] + p(B)[U(\$100) - U(\$0)]$$

or (more intuitively)

$$p(A)[U(\$100) - U(\$0)] - p(B)[U(\$420) - U(\$300)].$$

Note that we need only know the differences of the utilities in each column, not their values. We ask which difference matters more to us, which has a greater affect on the achievement of our goals. Note also that the probabilities matter. Since the first term is multiplied by $p(A)$, the higher $p(A)$, the more we favor option T over option S. This is a basic principle of decision making: options should be favored more when the probability of good outcomes (our degree of belief in them) is higher and the probability of bad outcomes is lower. This principle follows from the most basic assumptions about what decisions involve (e.g., Irwin, 1971).

This table can be used to illustrate an argument for EUT (Köbberling & Wakker (2001). As usual, the rows are acts, the columns are states, and the cells are outcomes. In Choice 1, Option S yield \$300 if event A happens (e.g., a coin comes up heads) and \$100 if B happens. Köbberling & Wakker (2001) consider patterns like those for Choices 1–4. Suppose you are indifferent between S and T in Choice 1, between U and V in choice 2, and between W and X in Choice 3. Then you ought to be indifferent between Y and Z in Choice 4. Why? Because rational indifference means that the reason for preferring T if A happens, the \$120 difference, is just balanced by the reason for preferring S if B happens. Thus, we can say that the difference between 300 and 420 in state B just offsets the difference between 0 and 100 in state A. If you decide in terms of overall good, then the

Choice 1	State	
Option	*A*	*B*
S	\$300	\$100
T	\$420	\$0

Choice 2	State	
Option	*A*	*B*
U	\$500	\$100
V	\$630	\$0

Choice 3	State	
Option	*A*	*B*
W	\$300	\$210
X	\$420	\$100

Choice 4	State	
Option	*A*	*B*
Y	\$500	\$210
Z	\$630	\$100

Table 2.1: Four choices illustrating tradeoff consistency

300–420 difference in A is just as good (on the whole, taking into account the probability of A) as the 0–100 difference in B. Similarly, if you are indifferent in Choice 2, then the 500–630 difference just offsets the same 0–100 difference. So the 500–630 difference is also just as good. And if you are indifferent in Choice 3, then the 500–630 difference in A just offsets the 100–210 difference in B. So all these differences are equal in terms of good. In this case, you ought to be indifferent in Choice 4, too.

This kind of "trade-off consistency," in which Choices 1–3 imply the result of Choice 4, plus a couple of other much simpler principles, *implies expected utility theory.* In particular, you can use one of the differences, like the 100–210 difference in B, as a measuring rod, to measure off equal intervals under A. Each of these differences represents the same utility difference. Note that this is all we need. We do not need to know what "zero utility" is, because decisions always involve comparison of options. (Even doing nothing is an option.) And the unit, like many units, is arbitrary. Once we define it, we must stick with it, but we can define it as we like. In this example, the 100–210 difference under B is a unit. If trade-off consistency failed, we would not be able to do this. The utility measure of some difference in state A would change depending on what we used as the unit of measurement.

Later I shall explain why this analysis also implies that we need to multiply by probability. It should be clear for now, though, that the conditions for EU are met if we multiply the utility difference in column A by the same number in all the tables, and likewise for column B.

Why should trade-off consistency apply? The critical idea here is that good (or bad) results from what happens, not from what does not happen. Thus, *the effect on goal achievement of changing from one outcome to another in State A (e.g., $300 to $420 in Table 2.1 cannot change as a function of the difference between the two outcomes in State B (e.g., $100 vs. $0 or $210 vs. $100), because the states are mutually exclusive.* This conclusion is the result of imposing an analytic scheme in which everything we value about an outcome is assigned to the cell in which that outcome occurs. If, for example, we experience emotions that result from comparing what happened to what did not happen, then the experience of those emotions must be considered part of the outcome in the cell representing what happened. (In cases like those in Table 2.1, this could mean that the outcome is not fully described by its monetary value, so that the same monetary value could be associated with different outcomes in different sub-tables.)

Note that we are also assuming that the idea of differences in utility is meaningful. But it must be meaningful if we are to make such choices at all. For example, if States A and B are equally likely, then any choice between S and T must depend on which difference is larger, the difference between the outcomes in A (which favor option T) or the difference between the outcomes in B (which favor S). It makes sense to say that the difference between 200 to 310 has as much of an effect on goodness as the difference between 0 and 100.

In sum, the justification of EUT is based on the idea that columns of the table have independent effects on goodness, because we analyze decisions so that all the relevant consequences of a given option in a given stats fall into a single cell. Consequences do not affect goodness when they do not occur. Once we assume this framework, then we can use the difference between consequences under one state as a measuring stick, to mark off units of utility in another state, and vice versa. We can assign utilities to outcomes in such a way that the option that does the most good on the whole is always a weighted sum, where each column has its own weight. (See Baron, 2000, for

related argument.) The next step is to show what this weight has to do with probability.

2.4 Probability

The idea of probability has a long history, although the idea that probability is relevant to decision making is only a few hundred years old (Hacking, 1975). Scholars have distinguished several different ways of thinking about what probability means. A standard classification distinguishes three general approaches: necessary (logical), objectivistic, and personal (Savage, 1954).

The logical view sees probability as an extension of logic, often by analysis of situations into possible worlds and their enumeration. It is the view that is often implicit in the early chapters of textbooks of probability and statistics, where probability is introduced in terms of gambling. There is a sense in which the probability of drawing a king from a deck of cards is necessarily 1/13. That is part of what we mean by a "fair deck." Similarly, the probability of drawing a red card is 1/2, and the probability of drawing a red king is 1/26.

The logical view is not very useful for calculation of insurance premiums or analysis of experiments. Thus, the later chapters of statistics books generally switch to the view of probabilities as objective, as relative frequencies. By this view, the probability of drawing a king from a deck is ultimately defined as the relative frequency of kings to draws with an infinite number of draws. Likewise, the probability that you will live to be 100 years old is to be determined by counting the number of people like you who did and who did not live that long.

Two problems arise with this view. One is that "like you" is definable in many ways, and the frequencies are different for different definitions. Another is that sometimes we like to talk about the probability of unique events, such as the probability that the Democratic Party will win the next U.S. presidential election — not just elections in general, but that one in particular. Now it may be that such talk is nonsense, but the personal view assumes that it is not. And this is not a fringe view. It is often called Bayesian because it was advanced in a famous essay of Thomas Bayes (1764/1958), although had earlier antecedents (Hacking, 1975). The idea of the personal

view is that probability is a measure of a person's degree of belief in the truth of propositions (statements that can have a truth value). Thus, two people can have different probabilities for the same proposition.

You might think that the personal view is so loose that anything goes. It is true that it does not assume a right answer about the probability of some proposition. But it does have some requirements, so it can serve as a normative model of probability judgments. Two sorts of requirements have been proposed: calibration and coherence.

Calibration is, in a way, a method for incorporating the objective view into the personalist view. But it solves the problem of multiple classification by classifying the judgments themselves. Thus, all of a given judge's judgments of the form, "The probability of X is .8" are put together. If we then discover the truth behind each judgment, we should expect that 80% of the propositions are true.

Coherence is the requirement that sets of judgments must obey certain rules. The basic rules that define the concept of coherence are the following:

- The probability of a proposition's being true, plus the probability of its being false (called the probability of the *complement* of the proposition), must equal 1. A probability of 1 represents certainty.

- Two propositions, A and B, are *mutually exclusive* if they cannot both be true at the same time. If you believe that A and B are mutually exclusive, then $p(A) + p(B) = p(A \text{ or } B)$: That is, the probability of the proposition "either A or B" is the sum of the two individual probabilities. If we assume that "It will rain" and "It will snow" are mutually exclusive propositions (that is, it cannot both rain and snow), then the probability of the proposition "It will rain or it will snow" is the sum of the probability of rain and the probability of snow. This rule is called *additivity*.

- A definition: The *conditional probability of proposition A given proposition B* is the probability that we would assign to A if we knew that B were true, that is, the probability of A

conditional on B being true. We write this as $p(A/B)$. For example, $p(\text{king/face card}) = 1/3$ for an ordinary deck of cards (in which the face cards are king, queen, and jack).

- The *multiplication rule* says that $p(A \,\&\, B) = p(A/B) \cdot p(B)$. Here $A \,\&\, B$ means "both A and B are true." For example, if we think that there is a .5 probability of a given person being female and that the probability of a female's being over 6 feet tall is .02, then our probability for the person being a female over 6 feet tall is $p(\text{tall} \,\&\, \text{female}) = p(\text{tall/female}) \cdot p(\text{female}) = (.02) \cdot (.5) = .01$.

- In a special case, A and B are *independent*. Two propositions are independent for you if you judge that learning about the truth or falsity of one of them will not change your degree of belief in the other one. For example, learning that a card is red will not change my belief that it is a king. In this case, we can say that $p(A/B) = p(A)$, since learning about B does not change our probability for A. The multiplication rule for independent propositions is thus $p(A \,\&\, B) = p(A) \cdot p(B)$, simply the product of the two probabilities. For example, $p(\text{king} \,\&\, \text{red}) = p(\text{king}) \cdot p(\text{red}) = (1/13) \cdot (1/2) = 1/26$.

Such rules put limits on the probability judgments that are justifiable. For example, it is unjustifiable to believe that the probability of rain is .2, the probability of snow is .3, and the probability of rain *or* snow is .8. If we make many different judgments at one time, or if our past judgments constrain our present judgments, these constraints can be very strong. These constraints do not determine a *unique* probability for any proposition, however. Reasonable people can still disagree.

The two main rules here are additivity and multiplication. The rule concerning complements is a special case of additivity, simply defining the probability of a true proposition as 1. And the independence rule is a special case of the multiplication rule for the case in which the conditional probability and the unconditional probability are the same. Notice that all the rules are here defined in terms of relations among beliefs (following von Winterfeldt & Edwards, 1986).

2.4.1 Coherence rules and expected utility

Why are these rules normative? You might imagine a less demanding definition of coherence, in which the only requirement is that stronger beliefs (those more likely to be true) should be given higher numbers. This would meet the most general goal of quantifying the strength of belief. Or, looking at it another way, you might imagine that some transformation of p would do just as well as p itself. Why not p^2, or $p + .5$. Such transformations would violate either the addition rule or the multiplication rule, or both.[4] A major argument for the two rules comes from the use of probability in decisions.

In the section on EUT, I argued that the states — represented by columns — had corresponding weighting factors, which multiplied the utilities of each outcome in each column. EUT says that these are probabilities, and it does make sense to give more weight to outcomes in states that are more likely to happen. This requirement as stated, however, implies only an ordinal concept of probability, one that allows us to rank beliefs for their strength. As stated so far, it does not imply that probabilities must follow the coherence rules.

To see how it actually does imply coherence, consider the possibility of re-describing decisions in ways that do not affect good (goal achievement). Such a re-description should not affect the conclusions of a normative model about what we should do. This principle, called "extensionality" (Arrow, 1982) or "invariance" (Tversky & Kahneman, 1986), is important in the justification of several forms of utility theory.

First consider the addition rule. We can subdivide any state into two states. For example, if a state is about the weather tomorrow, for a decision about what sort of outing to take, we could subdivide the state "sunny" into "sunny and this coin comes up heads" and "sunny and this coin comes up tails." Any normative theory should tell us that this subdivision should not affect our decision. It is clearly irrelevant to good, to the achievement of our goals. Yet, if probability is not

[4]The square does not violate the multiplication rule: $p^2q^2 = (pq)^2$. But it does violate the addition rule: in general $p^2 + q^2 \neq (p + q)^2$. Addition of a constant violates both rules.

additive, it could change our decision.

For example, suppose that $p(\mathrm{S})$ is .4, where S is "sunny." If we subdivide into sunny-heads and sunny-tails, the probability of each would be .2. Additivity applies, and our decision would not change. In particular, if P is "picnic," H is "heads," and T is tails, $p(\mathrm{S})U(\mathrm{PS}) = p(\mathrm{SH})U(\mathrm{PSH}) + p(\mathrm{ST})U(\mathrm{PST})$. The utilities here are all the same, of course. Now suppose we transform p so that additivity no longer applies. For example, we add .1 to each p. In this case, we would add .1 to the left side of the last equation (because there is one probability) and .2 to the right side (because there are two), and the equality would no longer hold. Of course this would not necessarily change our decision. The same option (e.g., movie, vs. picnic) might win. But it might not, because the calculated EU of picnic might increase. The only way to avoid such effects of arbitrary subdivision is to require that the addition rule apply to the weights used to multiply the states.

Now consider the multiplication rule. This relates to a different kind of invariance. Our choices should not be affected if we narrow down the space of possibilities to what matters. Nor should it matter what series of steps got us to the consequences, so long as the consequences represent fully everything that affects our good. (Again, this is a requirement of the analytic scheme that we impose.) The irrelevance of the series of steps is called "consequentialism" (Hammond, 1988), although that term has other meanings.

Consider the sequence of events in Figure 2.1, a classic case originally discussed by Allais (1953; and discussed further by Kahneman & Tversky, 1979; McClennan, 1990; and Haslam & Baron, 1993). You have a .25 probability of getting to point A. If you don't get there, you get \$0. If you get there, you have a choice between Top and Bottom, the two branches. Top gets you \$30 for sure. Bottom gets you a .8 probability of \$45, but if you don't get \$45 you get \$0.

You can think about the choice you must make from your point of view now, before you know whether you get to A. That situation is represented as follows:

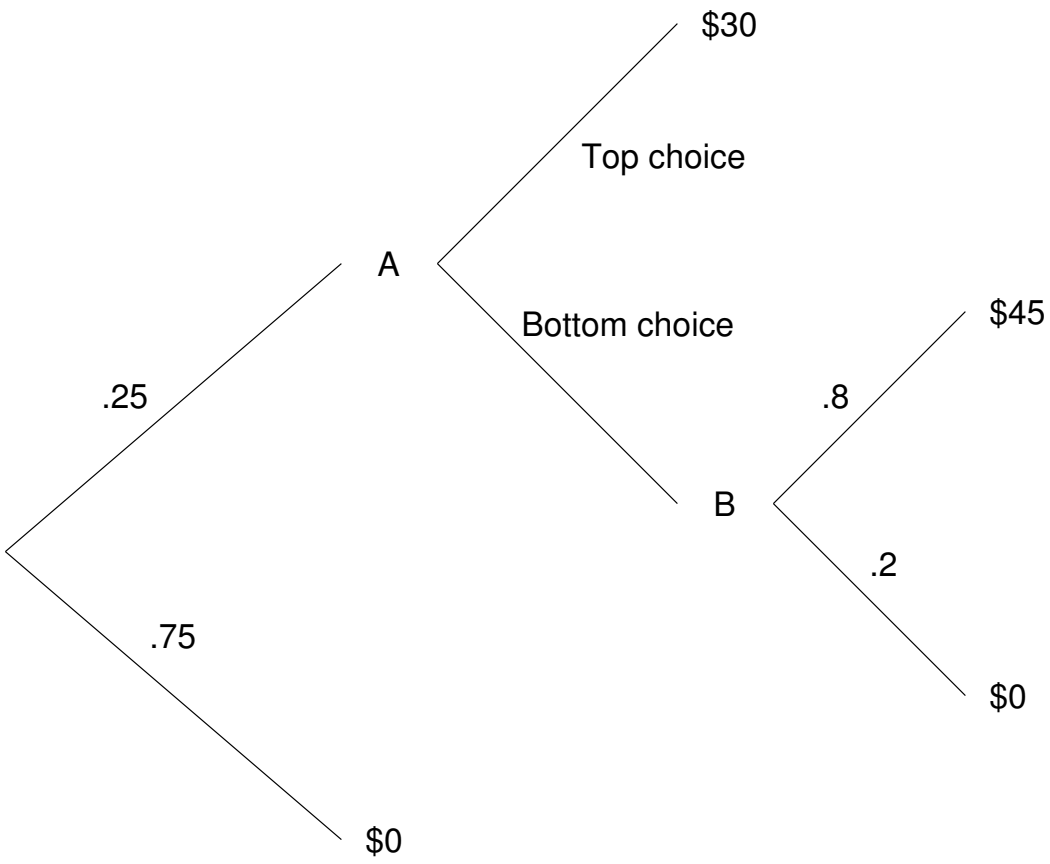

Figure 2.1: Illustration of consequentialism: A is a choice; B is a chance event.

Probability:	.75	.20	.05
Top	\$0	\$30	\$30
Bottom	\$0	\$45	\$0

Alternatively, you can think about your choice from your point of view if you get to A:

Probability:	.8	.2
Top	\$30	\$30
Bottom	\$45	\$0

If the money is all you care about — and we are assuming that it is, since this is just an example where the money stands for any consequences that are the same when they have the same labels — then the choice that is better for you (the one that achieves your goals better) should not depend on when you make it. It depends only on the outcomes. This is because the time of making the choice does not affect which option is better (achieves your goals more).

Notice how the probabilities of .20 and .05 come about. They are the result of the multiplication rule for probabilities. The .20 is the product $p(A)p(\$45/A)$, that is, the probability of A times the conditional probability of \$45 given A. If the multiplication rule did not hold, then the numbers in the top table would be different. The relative EU of the two options could change as well. In particular, the relative utility of the two options depends on the ratio of the probability of \$45 and the probability of \$0 in case Bottom is chosen. In sum, if the multiplication rule were violated, consequentialism could be violated.[5]

In sum, EUT requires that the basic coherence constraints on probability judgments be met. If they are not met, then different ways of looking at the same situation could lead to different conclusions about what to do. The situation is the "same" because everything that affects goodness (goal achievement) is the same, and this is a result of how we have analyzed the situation.

[5]Another way to look at this situation is to adopt the perspective of some scholars of probability, who say "all probabilities are conditional." Probabilities that seem not to be conditional are simply conditional on your current beliefs. The change from the perspective of the original decision maker to that of the person at point A simply involves a narrowing of the frame of reference. Again, this should not affect the relative goodness of the two options.

2.4.2 Bayes's theorem

The multiplication and additivity rules together imply a famous result, developed by Bayes, called Bayes's theorem. The result is a formula, which can be used for reversing a conditional probability. For example, if H is a hypothesis (the patient has a bacterial infection) and D is a datum (the throat culture is positive), we can infer $p(H/D)$ from $p(D/H)$ and other relevant probabilities. If we make judgments of both of these probabilities, we can use the formula to determine whether these judgments are coherent. The formula is a normative model for some judgments.

Specifically, we can calculate $p(H/D)$ from the multiplication rule,
$p(H \,\&\, D) = p(H/D) \cdot p(D)$, which implies

$$p(H/D) = \frac{p(H \,\&\, D)}{p(D)} \tag{2.1}$$

Formula 1 does not help us much, because we don't know $p(H \,\&\, D)$. But we do know $p(D/H)$ and $p(H)$, and we know from the multiplication rule that $p(D \,\&\, H) = p(D/H) \cdot p(H)$. Of course $p(D \,\&\, H)$ is the same as $p(H \,\&\, D)$, so we can replace $p(H \,\&\, D)$ in formula 1 to get:

$$p(H/D) = \frac{p(D/H) \cdot p(H)}{p(D)} \tag{2.2}$$

Formula 2 is useful because it refers directly to the information we have, except for $p(D)$. But we can calculate that too. There are two ways for D to occur; it can occur with H or without H (that is, with $\tilde{}H$). These are mutually exclusive, so we can apply the additivity rule to get:

$$\begin{aligned} p(D) &= p(D \,\&\, H) + p(D \,\&\, \tilde{}H) \\ &= p(D/H) \cdot p(H) + p(D/\tilde{}H) \cdot p(\tilde{}H) \end{aligned}$$

This leads (by substitution into formula 2) to formula 3:

$$p(H/D) = \frac{p(D/H) \cdot p(H)}{p(D/H) \cdot p(H) + p(D/\tilde{}H) \cdot p(\tilde{}H)} \tag{2.3}$$

Formulas 2 and 3 are called *Bayes's theorem*. In formula 3, $p(H/D)$ is usually called the *posterior probability* of H, meaning the probability after D is known, and $p(H)$ is called the *prior probability*, meaning the probability before D is known. $p(D/H)$ is sometimes called the *likelihood* of D.

2.5 Utilitarianism

Utilitarianism extends the basic normative model of EUT to people as well as states. So far I have been talking about "our goals" as if it didn't matter whether decisions were made for individuals or groups. But the best option for one person is not necessarily the best for another. Such conflict between people is analogous to the conflict that arises from different options being better in different states of the world. We can extend our basic table to three dimensions.

Utilitarianism, of course, is a traditional approach to the philosophy of morality and law, with antecedents in ancient philosophy, and developed by Bentham (1843/1948), Mill (1863), Sidgwick (1907/1962), Hare (1981), Broome (1991), Baron (1993), and Kaplow and Shavell (2002), among others. It holds that the best choice is the one that does the most good. This is just what utility theory says, in general, but utilitarianism requires the somewhat controversial claim that utility can be added up across people just as it can across uncertain states. In utilitarianism, "each person counts as one, no person as more than one," so the idea of weighting is absent.

Utilitarianism is, in fact, very closely related to EU theory. It is difficult to accept one and not the other. When a group of people all face the same decision, then the two models clearly dictate the same choice. For example, suppose that we have a choice of two policies for distribution of annual income among some group of people. Plan A says that half of them, chosen at random, will get \$80,000 per year and half will get \$40,000. Plan B gives everyone \$55,000. Suppose that, given each person's utility for money, plan A has a higher EU for each person (before it is known which group he is in). Utilitarianism would require the choice of plan A in this case: the total utility of A would necessarily be higher than B. The difference between EU and utilitarianism is that utilitarianism multiplies the average utility of \$80,000 by the number of people who get it,

etc., while EU multiplies the same average utility by each person's probability of getting it. It is easy to see that these calculations must yield the same result.

Suppose, on the other hand, we take a non-utilitarian approach to distribution. We might think that fairness requires a more equal distribution, at the cost of some loss in total utility, so we might choose B. In this case, viewed from the perspective of each person *ex ante* (before it is determined who is in which group), *everyone* is worse off than under plan A.

Kaplow and Shavell (2002) show how *any* non-utilitarian principle of distribution can yield results like this, in which nobody is better off and at least some people are worse off. The fact that these situations are highly hypothetical is not relevant, for a *normative* theory should apply everywhere, even to hypothetical situations. Of course, this argument, by taking an *ex-ante* perspective, assumes that people's good can be represented by EUT. But that is the point; EUT and utilitarianism are intimately connected.

Broome (1991) argues more formally that, if the utility of each person is defined by EU, then, with some very simple assumptions, the overall good of everyone is a sum of individual utilities. In particular, we need to assume the "Principle of Personal Good" (Broome, 1991, p. 165): "(a) Two alternatives are equally good if they are equally good for each person. And (b) if one alternative is at least as good as another for everyone and definitely better for someone, it is better." This is a variant of the basic idea of "Pareto optimality." This variant assumes that the probabilities of the states do not depend on the person. It is as though the theory were designed for decision makers with their own probabilities.

To make this argument, Broome (1991) considers a table like the following. Each column represents a state of nature (like those in the expected-utility table above). Each row represents a person. There are s states and h people. The entries in each cell represent the utilities for the outcome for each person in each state. For example u_{12} is the utility for Person 1 in State 2.

$$\begin{array}{cccc} u_{11} & u_{12} & \ldots & u_{1s} \\ u_{21} & u_{22} & \ldots & u_{2s} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ u_{h1} & u_{h2} & \ldots & u_{hs} \end{array}$$

The basic argument, in very broad outline, shows that total utility is an increasing function of both the row and column utilities, and that the function is additive for both rows and columns. Expected-utility theory implies additivity for each row. For a given column, the Principle of Personal Good implies that the utility for each column is an increasing function of the utilities for the individuals in that column. An increase in one entry in the cell has to have the same effect on both the rows and the columns, so the columns must be additive too. [6]

2.6 Conclusion

I have sketched some of the arguments for some of the major normative models used in JDM. I have emphasized one approach, based on the idea that models are justified by the imposition of an analytic framework, based on very fundamental assumptions about the nature of humans.

The importance of the analytic framework may be illustrated in terms of the way it handles apparent counter-examples. Consider the following three choices offered (on different days) to a well-mannered person (based on Petit, 1991):

1. Here is a (large) apple and an orange. Take your pick; I will have the other.

2. Here is an orange and a (small) apple. Take your pick; I will have the other.

3. Here is a large apple and a small apple. Take your pick; I will have the other.

It would make sense to choose the large apple in Choice 1 and the orange in Choice 2. But a polite person would choose the small apple in Choice 3, thus (apparently) violating transitivity. It is

[6] For details, see Broome, 1991, particularly, pp. 68, 69, and 202.

impolite to choose the large apple when the only difference is size, but it is acceptable to choose the larger fruit when size is not the only difference. But Choice 3 is not just between a large apple and a small apple. It is between "a large apple plus being impolite" and a small apple. The impoliteness associated with the large apple reduces its utility and makes it less attractive. Transitivity is not actually violated. When we use utility theory to analyze decisions, we must make sure to include all the relevant consequences, not just those that correspond to material objects.

For the sake of brevity, I have omitted some extensions of utility theory that are important in JDM. One important extension is multi-attribute utility theory (chs. 16 and 17), which applies when the "parts" of utility are attributes such as the price, speed, and memory size of a computer. The attributes must have independent effects on utility for the simplest version of this model to be relevant, so it requires proper analysis, which is not always possible. This model is discussed by Keeney and Raiffa's classic work (1976/1993) and by Keeney (1992), among other works.

A second model concerns choice over time, when outcomes occur at different times (ch. 21). A different form of the idea of dynamic consistency is often applied to this situation: the decision should not change as a function of when it is made, so long as the outcomes are not affected (Baron, 2000, ch. 19).

Finally, I have omitted some of the more intricate arguments that have occupied scholars over the last few decades (e.g., Bachrach & Hurley, 1991). It is my hope that, in the long view of history, most of these arguments will be seen as necessary for the purpose of arriving at a good theory, but ultimately irrelevant once the theory is refined and widely understood. The idea of utility theory is simple — do the most good — and its simplicity, like that of arithmetic, may be what lasts.

2.7 References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'école américaine. *Econometrica, 21*, 503–546.

Arkes, H. R. (1996). The psychology of waste. *Journal of Behavioral Decision Making, 9*, 213–224.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 124–140.

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Inquiry, 20*, 1–9.

Bachrach, M. O. L. & Hurley, S. L. (Eds.) *Foundations of decision theory: issues and advances.* Oxford: Blackwell.

Baron, J. (1985). *Rationality and intelligence.* New York: Cambridge University Press.

Baron, J. (1993). *Morality and rational choice.* Dordrecht: Kluwer.

Baron, J. (1994). Nonconsequentialist decisions (with commentary and reply). *Behavioral and Brain Sciences, 17*, 1–42.

Baron, J. (1996). Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy, 12*, 165–182.

Baron, J. (2000). *Thinking and deciding* (3d ed.). New York: Cambridge University Press.

Baron, J. (2002). Value trade-offs and the nature of utility: bias, inconsistency, protected values, and other problems. Paper for conference on behavioral economics. American Institute for Economic Research, Great Barrington, MA, July, 2002.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika, 45*, 293–315. (Original work published 1764)

Bentham, J. (1948). *An introduction to the principles of morals and legislation.* Oxford: Blackwell Publisher. (Original work published 1843)

Broome, J. (1991). *Weighing goods: Equality, uncertainty and time.* Oxford: Basil Blackwell.

Broome, J. (1997). Is incommensurability vagueness? In R. Chang (Ed.), *Incommensurability, incomparability, and practical reason*, pp. 67–89. Cambridge, Mass.: Harvard University Press.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

de Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. (Translated by H. E. Kyburg, Jr., and H. E. Smokler.) In H. E. Kyburg, Jr., and H. E. Smokler (Eds.) *Studies in subjective probability.* New York: Wiley, 1964.

Hacking, I. (1975). *The emergence of probability.* New York: Cambridge University Press.

Hammond, P. H. (1988). Consequentialist foundations for expected utility. *Theory and decision, 25*, 25–78.

Hare, R. M. (1981). *Moral thinking: Its levels, method and point.* Oxford: Oxford University Press (Clarendon Press).

Haslam, N., & Baron, J. (1993). Rationality and resoluteness: Review of *Rationality and dynamic choice: Foundational Explorations*, by E. F. McClennan. *Journal of Mathematical Psychology 37*, 143–153.

Irwin, F. W. (1971). *Intentional behavior and motivation: A cognitive theory.* Philadelphia: Lippincott.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kaplow, L., & Shavell, S. (2002). *Fairness versus welfare.* Cambridge, MA: Harvard University Press.

Keeney, R. L. (1992). *Value-focused thinking: A path to creative decisionmaking.* Cambridge, MA: Harvard University Press.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preference and value tradeoffs.* New York: Cambridge University Press. (Originally published, 1976.)

Köbberling, V. & Wakker, P. P. (2001). *A tool for qualitatively testing, quantitatively measuring, and normatively justifying expected utility.* Department of Quantitative Economics, University of Maastricht, Maastricht, The Netherlands.

`http://www.fee.uva.nl/creed/wakker/newps.htm`

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol.

1). New York: Academic Press.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science, 1*, 362–370.

McClennan, E. F. (1990). *Rationality and dynamic choice: Foundational Explorations.* New York: Cambridge University Press.

Mill, J. S. (1863). *Utilitarianism.* London: Collins.

Petit, P. (1991). Decision theory and folk psychology. In M. O. L. Bachrach & S. L. Hurley (Eds.) *Foundations of decision theory: issues and advances*, pp. 147–167. Oxford: Blackwell.

Popper, K. R. (1962). Why are the calculi of logic and arithmetic applicable to reality? Chapter 9 in *Conjectures and refutations: The growth of scientific knowledge*, pp. 201–214. New York: Basic Books.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays by F. P. Ramsey* (pp. 158–198). New York: Harcourt, Brace.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Harvard University Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Sidgwick, H. (1962). *The methods of ethics. (7th ed., 1907)* Chicago: University of Chicago Press. (First edition published 1874).

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*, S251–S278.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton: Princeton University Press.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* New York: Cambridge University Press.

Wakker, P. (1989). *Additive representation of preferences: A new foundation of decision analysis.* Dordrecht: Kluwer.