

1

Instrumental Rationality as Expected Utility Maximization

This chapter provides a brief quasi-historical introduction to *expected utility theory*, the most widely defended version of *normative decision theory*. The overarching goal of normative decision theory is to establish a general standard of rationality for the sort of *instrumental* (or “practical”) reasoning that people employ when trying to choose means appropriate for achieving ends they desire. Expected utility theory champions *subjective expected utility maximization* as the hallmark of rationality in this means-ends sense.

We will examine the theory in the setting where it works best by applying it to the case of professional gamblers playing games of chance inside casinos. In this highly idealized situation, the end is always the maximization of one’s own fortune, and the means is the ability to buy and sell wagers that offer monetary payoffs at known odds. Later chapters will consider more general contexts. Since the material here is presented in an elementary (and somewhat pedantic) way, those who already understand the concept of expected utility maximization and the rudiments of decision theory are encouraged to proceed directly to Chapter 2.

1.1 PASCAL AND THE “PROBLEM OF THE POINTS”

The *Port Royal Logic* of 1662 contains the first general statement of the central dogma of contemporary decision theory:

In order to decide what we ought to do to obtain some good or avoid some harm, it is necessary to consider not only the good or harm in itself, but also the probability that it will or will not occur; and to view geometrically the proportion that all these things have when taken together.¹

In modern terms, the suggestion here is that risky or uncertain prospects are best evaluated according to the *principle of mathematical expectation*, so that “our fear of some harm [or hope of some good] ought to be proportional not only to the magnitude of the harm [or good], but also to [its] probability.”²

This principle, and the theory of probability that underlies it, had been discovered in 1654 by Blaise Pascal, the greatest of the many great thinkers that Port Royal produced, during the course of a correspondence with Pierre de Fermat concerning a gambler’s quandary now known as the *problem of the points*.³ It had been posed to Pascal by a “reputed gamester,” the Chevalier de Méré, who Pascal regarded as a fine fellow even

though he suffered from the “great fault” of not being a mathematician. The question had to do with the fair division of a fixed pot of money among gamblers forced to abandon a winner-take-all game before anyone had won. Here is a simplified version of the problem that Pascal and Fermat considered (with dollars instead of “pistoles” as currency): Two gamblers, H and T, are playing a game in which a coin, known to be fair, is to be tossed five times and a \$64 prize awarded to H or T depending on whether more heads or tails come up. Suppose that the first three tosses go head/tail/tail, and that the game is then interrupted, leaving the two gamblers with the task of finding an equitable way to dividing the \$64. T, who has two of the three tails she needs to win, would surely feel cheated if the pot were split down the middle. H, on the other hand, would be justifiably upset if T got all the money since he still had a chance to win the game when it was stopped. Clearly, the fair division must give T something more than \$32 but less than \$64. The challenge for Pascal and Fermat was to find the right amount. Both men solved the instance of the problem of the points they were considering, but Pascal, in a great feat of mathematical genius, went on to treat the general case.

His solution had two parts. First, he invented the theory of probability more or less from scratch. Professional gamblers had long known that one could use nonnegative real numbers to measure the frequencies at which various events occur, and that these would give the odds at which various bets would be advantageous. Legend has it, for example, that the Chevalier de Méré made a lot of money laying even odds that he could roll at least one 6 in four tosses of a fair die. What the Chevalier realized, and his gullible opponents did not, was that the probability of this event was slightly more than one-half (about 0.518), and thus he was likely to win his bet more often than not. Unfortunately, probabilities were difficult to calculate, and gamblers were forced to find them empirically by observing the relative frequencies at which various events occurred. This method worked well for simple events, but it was hard to apply in even modestly complicated cases. In fact, after people had seen through his first scam, de Mere nearly bankrupted himself by overestimating his chances of throwing a 12 in twenty-four tosses of a pair of fair dice. Pascal solved the gamblers’ problem by discovering

The Fundamental Law of Probability. If $\{E_1, E_2, \dots, E_n\}$ is any set of jointly exhaustive, mutually exclusive events, each of which has a definite probability, then the sum of all these probabilities is 1, that is, $\sum_j \rho(E_j) = \rho(E_1) + \rho(E_2) + \dots + \rho(E_n) = 1$.

While he never expressed it quite so explicitly, there is no doubt that Pascal understood this principle and recognized it as the key to calculating probabilities. He would also have endorsed the modern equivalent reformulation of his law in terms of the following three *axioms of probability*, which are supposed to hold for all events E and E^* (where \neg and \vee are the Boolean operations of negation and disjunction):

Nonnegativity. $\rho(E) \geq 0$

Normalization. $\rho(E \vee \neg E) = 1$

Finite Additivity. $\rho(E \vee E^*) = \rho(E) + \rho(E^*)$, whenever E and E^* are mutually incompatible events

He may also have endorsed the further axiom (which is of somewhat more recent vintage)

Continuity. For any denumerably infinite sequence of events E_1, E_2, E_3, \dots whose probabilities are well defined one has

$$\lim_{n \rightarrow \infty} \rho(E_1 \vee E_2 \vee \dots \vee E_n) = \rho(E_1 \vee E_2 \vee E_3 \vee \dots)$$

Note for future reference that Continuity and Finite Additivity ensure that probabilities are *countably additive* in the sense that $\sum_j \rho(E_j) = \rho(E_1 \vee E_2 \vee \dots)$ whenever the E_j form a countable sequence of mutually incompatible events.

Pascal brought his new theory of probabilities to bear on games of chance by noting that any listing of the possible endings to a gambling game is always a *partition* of mutually exclusive, collectively exhaustive events. For the version of the problem of the points considered here, one partition of endgames (in obvious notation) is $\{E_{HH}, E_{HT}, E_{TH}, E_{TT}\}$, another is $\{(E_{HH}, (E_{HT} \vee E_{TH}), E_{TT})\}$, and a third is $\{E_{HH}, (E_{HT} \vee E_{TH} \vee E_{TT})\}$. Applied to these three cases, Pascal's basic insight was that

$$\begin{aligned} 1 &= \rho(E_{HH}) + \rho(E_{HT}) + \rho(E_{TH}) + \rho(E_{TT}) \\ &= \rho(E_{HH}) + \rho(E_{HT} \vee E_{TH}) + \rho(E_{TT}) \\ &= \rho(E_{HH}) + \rho(E_{HT} \vee E_{TH} \vee E_{TT}) \end{aligned}$$

Since the coin is fair $1/4 = \rho(E_{HH}) = \rho(E_{HT}) = \rho(E_{TH}) = \rho(E_{TT})$, and this determines the probabilities for all disjunctions involving these four events. For example, the probability that T had of winning the game had it been finished can be computed as $\rho(E_{HT} \vee E_{TH} \vee E_{TT}) = \rho(E_{HT}) + \rho(E_{TH}) + \rho(E_{TT}) = 3/4$.

Having thus invented probability theory, Pascal went on to propose that a player's fair share of the pot in a truncated game of points should always be her *expected payoff*, the quantity obtained by combining her winnings under all endgames "geometrically" (i.e., multiplicatively) with their probabilities. This means that a player who stood to win $\$x_i$ if the game had ended in the way described by E_i would have a claim on $\rho(E_i)x_i$ from the pot, and her total fair share could be found by summing over all possible endgames to get her expected payoff $\text{Exp}_s = \sum_i \rho(E_i)x_i$. In our example, T, who wins if a tail is thrown on either of the final two tosses, is awarded

$$[\rho(E_{HH}) \times 0 + \rho(E_{HT}) \times 64 + \rho(E_{TH}) \times 64 + \rho(E_{TT}) \times 64] = \$48$$

while H, who needs two heads to win, is entitled to

$$[\rho(E_{HH}) \times 64 + \rho(E_{HT}) \times 0 + \rho(E_{TH}) \times 0 + \rho(E_{TT}) \times 0] = \$16$$

Pascal’s solution to the problem of the points has many attractive features. First, since the expected payoffs for a group of players always add up to the amount in the pot, nothing ever goes to waste and no one is ever promised money that cannot be paid. Expected payoffs also have the right kind of symmetry properties for a “fair division” rule. It seems a minimum requirement of fairness that two players who had exactly the same chances of winning the same prizes when the game ended should be assigned the same share of the pot. The expected payoff scheme does this. Moreover, a division rule should never give one player a larger payoff than another if the second stood to win more money in *every* possible endgame. Indeed, if the second player were sure to win \$y more than the first in every case, then it would seem that her payoff should be exactly \$y larger. Again, expected payoffs have this desirable property. Finally, on Pascal’s proposal, a player’s fair share of the pot depends *only* on the monetary prizes she had a chance of winning when the game ended and the probabilities with which she might have won them.

This last point is particularly important because it entails that the only thing relevant to determining a bettor’s share of the pot is the *wager* that she happens to hold when the game ends, where a “wager” is just a specification of her potential winnings and the circumstances under which she stands to win them. In our example, the wager T faced just before the game ended was

Event:	Head/Head	Head/Tail	Tail/Head	Tail/Tail
Probability:	1/4	1/4	1/4	1/4
Outcome:	T wins \$0	T wins \$64	T wins \$64	T wins \$64

Notice that this makes no reference to H’s payoffs, to the overall size of the pot, or even to the history of the game. In any betting arrangement – no matter how many players are involved, no matter how much money is in the pot, no matter what kind of rules are in force – if T ends up in the situation described by the table then she is entitled to exactly \$48 under Pascal’s proposal.

One way to put this is to say that Pascal sets the *fair price* of the wager that T holds when the game ends at \$48. This is the sum of money at which T should be equally happy to have either a straight payment of the sum or the wager itself. In general, a wager *G*’s *fair price* for an agent is that amount of money \$g at which she would be indifferent between holding *G* and having a straight payoff of \$g. This price is characterized by the following two conditions:

- (a) The agent finds the prospect of having any fortune greater than \$g *strictly more desirable* than the prospect of holding *G*.

(b) She finds the prospect of having any fortune less than G *strictly less desirable* than the prospect of holding G .

Economists sometimes refer to G 's fair price as its "certainty equivalent" to indicate that it is a "riskless asset" that is equivalent in value to the "risky" prospect G .

Pascal's solution to the problem of the points thus has two parts: First, he claims that any bettor's fair share of the pot is the fair price of the wager she holds when the game ends. This makes sense given that the latter quantity is exactly what she could reasonably demand in compensation for ending the game voluntarily at that time. Next, he maintains that a wager's fair price should be identified with its expected payoff, the amount obtained by multiplying each of its prizes by its probability and then summing over the whole. The real substance of Pascal's proposal, then, is that *wagers should be valued according to their expected payoffs*.⁴ To appreciate the implications of this idea we shall need to spend a moment getting clearer about the concept of a wager and the notion of an expected payoff.

1.2 WAGERS, PROBABILITIES, AND EXPECTATIONS

Pascal seems to have thought of wagers as arrangements under which a gambler is sure to end up having one of a finite set of (positive or negative) sums $\$x_1, \$x_2, \dots, \$x_n$ *added* to her net worth depending upon which member of a partition of mutually exclusive, jointly exhaustive events E_1, E_2, \dots, E_n happens to occur. Here is an example in which the event partition lists all of the possible ways in which a fair die might fall

Die Falls:	1	2	3	4	5	6
Incremental Payoff:	\$11	-\$6	-\$13	-\$9	-\$5	\$66

Under this arrangement, the bettor would have \$11 added to her bank account if the die were to come up 1, \$6 would be subtracted if it were to come up 2, \$13 would be subtracted if were to come up 3, and so on.

We will conceive of wagers somewhat differently. First, rather than portraying payoffs as *changes* in wealth we will construe them as specifications of *levels of total wealth*. Thus, for a person whose current net worth is \$10,000 the correct description of the above wager would be

Die Falls:	1	2	3	4	5	6
Fortune:	\$10,011	\$9,994	\$9,987	\$9,991	\$9,995	\$10,066

while for someone with a net worth of \$25,000 it would be

Die Falls:	1	2	3	4	5	6
Fortune:	\$25,011	\$24,994	\$24,987	\$24,991	\$24,995	\$26,066

This may take some getting used to. Psychological research has show that people typically assess economic prospects in terms of their *incremental* effects on wealth rather than their effects on total wealth.⁵ We must be careful not to think this way, however, because we need to allow for a possibility that Pascal never envisioned, namely, that the attractiveness of a wager might depend on both the changes it can make to a gambler's fortune and the size of the fortune she starts out with. Indeed, we shall shortly see that the idea of valuing wagers by their expected payoffs founders on precisely this point.

We will also go beyond Pascal by thinking of every wager as being defined over *the same* partition of events. Following the classic treatment in Leonard Savage's *Foundations of Statistics*, we imagine that there is a partition \mathcal{S} of mutually exclusive, jointly exhaustive *states of the world* each of which specifies a determinate result for every contingency that could conceivably affect a bettor's fortune. So, for every coin ever tossed, every die ever thrown, every roulette wheel ever spun, every poker hand ever dealt, a state will say how the coin falls, how the die lands, where the wheel comes to rest, what cards are dealt, what cards a player would draw were he to draw them, and so on, and so on. While Savage left the notion of a state an unanalyzed primitive in his theory, we will follow the lead of Richard Jeffrey and construe states as *propositions*. I will say more about what this means in the next chapter, but it suffices for now to think of state propositions as *descriptions* of possible states of affairs. These descriptions can properly be called true or false, can be combined using Boolean connectives and other propositional operations, and can be used to represent the ultimate objects of the bettor's beliefs. Every state proposition will describe a possible course of events that is *complete* in the sense that it specifies a unique, determinate result for every wager that anyone might have occasion to consider. Thus, the states in \mathcal{S} should be so finely specified that a bettor's uncertainty about the outcomes of the wagers she holds can always be modeled as uncertainty about which state actually obtains.

Events are less specific propositions about possible world histories that can be expressed as disjunctions of states. For technical convenience, we will suppose that the set of all events or *event space*, here denoted $\Omega(\mathcal{S})$, is a σ -complete Boolean algebra. This is a set of propositions that is closed under negation and countable disjunction, so that $\neg E_j$ and $\bigvee_j E_j = (E_1 \vee E_2 \vee E_3 \vee \dots)$ are events whenever $E_1, E_2, E_3 \dots$ are events.

A wager, for current purposes, is simply a proposition that describes a way in which a bettor's total fortune depends on the state of the world. Such a proposition can be expressed as a conjunction of subjunctive conditionals of the form⁶

$$\begin{aligned}
 G &= \&_S(\text{If } S \text{ were true, then the gambler's total fortune would be } \$g_S) \\
 &= \&_S(S \square \rightarrow \$g_S)
 \end{aligned}$$

where S ranges over states of the world, and “ $\square \rightarrow$ ” stands for the subjunctive conditional “If ___ were the case, then ---- would also be the case.”

Economists make a distinction between two kinds of wagers or, better, between two states of a gambler’s knowledge about the events on which they depend. When a wager involves *risk* each state S has a definite objective probability $\rho(S)$ that is known to the bettor. In contrast, *uncertain* prospects are those in which the bettor is in the dark about the objective probabilities of at least some of the events on which her fortunes depend, either because she does not have access to them or because there are no such probabilities to be known. The paradigm example of an uncertain prospect is a bet on the finish of a horse race (described here for an agent with an initial fortune of \$10,000)

Events:	Stewball wins.	Stewball places.	Stewball shows.	Stewball runs in the field.
Payoffs:	\$10,012	\$10,006	\$10,003	\$10,002

Unlike the other cases we have considered, here it is unrealistic to expect an agent to know the probabilities of the events that will determine her fortune since the factors that decide horse races are too numerous and unpredictable to make knowledge of objective chances possible (even if they do exist).

Pascal was concerned, at least initially, with finding fair prices for wagers involving risk rather than uncertainty. When considering the problem of the points, for example, he assumed that each endgame would have a definite probability that would be known to all players. We will see that Pascal did consider extending his approach to broader contexts, but for now it will be useful for us to think of him as offering a theory of fair pricing for wagers involving risk. We shall thus assume that there is a function ρ that assigns each event E in the algebra $\Omega(\mathcal{S})$ an *objective probability* $\rho(E)$ between 0 and 1, and that our gamblers all know exactly what these probabilities are.

Once we have an objective probability function in hand we can define *objective expected values* for many functions that assign real numbers to states in \mathcal{S} . A real-valued function F of the states in \mathcal{S} is said to be *measurable* with respect to ρ if and only if the disjunctive proposition $F_x = \bigvee \{S \in \mathcal{S} : F(S) = x\}$ is in ρ ’s domain $\Omega(\mathcal{S})$ for each real x . In other words, F is measurable exactly if ρ assigns a definite probability to every proposition of the form $F_x =$ “ F ’s value is x .” When F is measurable it makes mathematical sense to define its ρ -*expectation* as $\text{Exp}_\rho(F) = \int_{\mathcal{S}} F(S) d\rho$. This integral is nothing more than a weighted average of F ’s values where the weight assigned to each value is the probability of the state that generates it.⁷ In the cases that will be of most interest to us F ’s *support*, the set of values whose associated F_x propositions have nonzero probability, will be at most countably infinite and the preceding expression will reduce to a straight sum of those values of F that have a nonzero probability of being realized weighted by their probabilities, so that $\text{Exp}_\rho(F) = \sum_x x\rho(F_x)$ where x ranges over the values in F ’s support.

Except in rare cases, which will be clearly noted, the reader will never go wrong thinking of expectations as sums of this sort. In what follows I will largely ignore the difference between integrals and sums, and will write $\text{Exp}_\rho(F) = \sum_S F(S)\rho(S)$. Strictly speaking, this identity will only hold when there are countably many states, but everything I will have to say about such sums will carry over to integrals of the form $\int_S F(S)d\rho$.

For future reference, note that expectations obey the following principles:

Constancy. If $F(S) = 1$ for all states S , then $\text{Exp}_\rho(F) = 1$.

Dominance. If $F(S) \geq F^*(S)$ for all states S , then $\text{Exp}_\rho(F) \geq \text{Exp}_\rho(F^*)$, and this inequality is strict when there is a nonzero probability that F 's value will be larger than F^* 's.

Additivity. $\text{Exp}_\rho(F + F^*) = \text{Exp}_\rho(F) + \text{Exp}_\rho(F^*)$ provided that the latter two expectations are defined.

Continuity. If $\text{Exp}_\rho(F_j)$ is defined for each $j = 1, 2, 3, \dots$, and if the sequence of functions F_1, F_2, F_3, \dots converges to F , then $\lim_{j \rightarrow \infty} \text{Exp}_\rho(F_j)$ converges and is equal to $\text{Exp}_\rho(F)$.

It is straightforward to show that these imply that the expectation operator is *linear*, so that $\text{Exp}_\rho(cF + c^*F^*) = c\text{Exp}_\rho(F) + c^*\text{Exp}_\rho(F^*)$ for any real numbers c and c^* . They also entail that expectations are *countably additive* in the sense that $\text{Exp}_\rho(\sum_j F_j) = \sum_j \text{Exp}_\rho(F_j)$, provided that all estimates are well defined and the right-hand sum converges.

There is a tight connection between wagers and real functions defined over states of the world. Each wager $G = \&_S(S \mapsto \$g_S)$ uniquely picks out a function $G(S) = g_S$, and any such function G seems to specify an associated wager $G = \&_S(S \mapsto \$G(S))$. Indeed, the connection is so close that many decision theorists *identify* wagers with abstract state-to-outcome functions.⁸ I prefer not to do this both because I want to insist on a conceptual distinction between the propositions that describe these mappings and the mappings themselves, and because I shall eventually reject the notion that every mapping from states to outcomes has a well-defined wager associated with it. The converse, however, is true, and we can use it to define a wager's expected payoff as follows:

Definition. If $G = \&_S(S \mapsto \$g_S)$ is a wager and $G(S)$ is its associated state-to-payoff function, then G 's *expected payoff*, or *actuarial value*, relative to the probability function ρ is $\text{Exp}_\rho(G) = \sum_{S \in S} G(S)\rho(S)$ (or $\int_S G(S)d\rho$ when the set of states is uncountable).

With this definition in hand we are in a position to give precise statement of Pascal's fair pricing scheme for risky wagers.

Pascal's Thesis. If a gambler believes that ρ is the correct objective probability function for states of the world, and if she is interested in the wager G solely as a means of

increasing her total fortune, then she should find the prospect of holding G and the prospect of having a sure fortune equal to G 's objective expected payoff equally desirable, so that for her $\$g = \text{Exp}_p(G)$

I have included the proviso that the bettor should be interested in G "solely as a means of increasing her fortune" because, as I think Pascal recognized, his thesis is only plausible when applied to "professional" gamblers who value nothing but money (at least when inside the casino), and who derive neither pleasure nor displeasure from the act of betting itself. Such a person will care about a wager only insofar as its truth might effect the size of her fortune. One way to put this is to say that she will regard money as *an unalloyed good* in the sense that she will find the prospects $(\$x \ \& \ E)$ and $(\$x \ \& \ \neg E)$ equally desirable for any event E and any wealth level $\$x$.

There are no "professional" bettors of course; people value money for what it can buy, not as an end in itself. This would be a serious problem if we were interested in applying Pascal's Thesis to real human beings, but this is not the plan. Our aim, instead, will be to see what the thesis tells us about the highly idealized, and thus more tractable, case of professional gamblers in the hopes of discovering general principles of instrumental rationality that can be applied in the more realistic settings to be considered in later chapters.

When combined with the basic properties of expectations, Pascal's Thesis imposes a number of restrictions on a rational professional bettor's fair prices.

Constancy. A wager $G = \&_S(S \ \square \rightarrow \ \$c)$ that is sure to leave a bettor with a fortune of $\$c$ in every state of the world has $\$c$ as its fair price.

Dominance. If $G = \&_S(S \ \square \rightarrow \ \$g_S)$ is sure to award at least as large a fortune as $H = \&_S(S \ \square \rightarrow \ \$h_S)$ does, so that $g_S \geq h_S$ for all $S \in \mathcal{S}$, then G 's fair price will be at least as great as H 's. Moreover, if there is a positive probability that G will produce a fortune strictly larger than H 's, then G 's fair price will be strictly higher than H 's.

Additivity. If $G = \&_S(S \ \square \rightarrow \ \$g_S)$ and $G^* = \&_S(S \ \square \rightarrow \ \$g_S^*)$ have fair prices of $\$g$ and $\$g^*$, respectively, then the fair price of the wager $H = \&_S(S \ \square \rightarrow \ \$(g_S + g_S^*))$ is $\$h = \$(g + g^*)$.

Continuity.: Let $g_S(1), g_S(2), g_S(3), \dots$ be real numbers converging to g_S for each $S \in \mathcal{S}$. Then, the fair price of $G = \&_S(S \ \square \rightarrow \ \$g_S)$ is the limit of the fair prices of the wagers $G(j) = \&_S(S \ \square \rightarrow \ \$g_S(j))$.

We will be evaluating each of these principles. But before doing so we need to correct a misconception about the nature of fair prices that has caused a great deal of confusion about the status of requirements of the sort just presented.

1.3 A SHORT DIGRESSION: AGAINST BEHAVIORISM

I was careful to formulate Pascal's Thesis as *a norm of rational desire* that governs the fair pricing of risky wagers, where fair prices are understood in terms of the pattern of preferences described in (a)-(b) of Section 1.1. I framed the issue this way because I hoped to distinguish my version of Pascal's Thesis from a closely related principle that governs not desires, but *overt actions*. In the economics literature one often finds fair prices defined not in terms of desires, but in terms of *overt choice behavior*: Instead of (a)-(b), the conditions for a gambler's having \$g as her fair price for G become

(a*) If asked to *choose* between holding G and having a guaranteed fortune of any sum greater than \$g, the gambler would choose the guaranteed fortune (provided that she is certain that nothing else of value hangs on her choice).

(b*) If asked to *choose* between holding G and having a guaranteed fortune of any sum less than \$g, the gambler would choose the wager (provided that she is certain that nothing else of value hangs on her choice).

This leads to the following revision:

Pascal's Thesis (as a constraint on acts). If a gambler believes that ρ is the correct objective probability function for states of the world, and if she is interested in a wager G only as a means of increasing her fortune, then she should be willing to exchange G to attain any wealth level that is not less than $\text{Exp}_\rho(G)$, and she should be willing to exchange any wealth level that is not greater than $\text{Exp}_\rho(G)$ to obtain G.

The two versions of Pascal's Thesis are closely related – anyone who thinks that gamblers should obey the first will agree that, *insofar as they are going to be in the business of buying and selling wagers*, they should obey the second. Nevertheless the two principles are conceptually distinct and the notion of a fair price defined in (a)-(b) is not the same as that defined in (a*)-(b*). In saying this I am departing from what has been the party line among rational choice theorists for many years. Decision theory came of age during the heyday of logical positivism, and many of its early practitioners embraced a kind of behaviorism that equated desires with *dispositions toward overt action*. They thus took it as *analytic* that an agent would have desires satisfying (a)-(b) if and only if she also had the behavioral dispositions described in (a*)-(b*).

It is easy to see why such a view might appeal to psychologists, economists, market researchers, or bookies interested in making an empirical study of the prices people pay for wagers. The challenge for these investigators is to find ways of making reliable inferences about the “internal” states and processes implicated in decision making on the basis of empirical information about patterns of overt behavior. Behaviorists solve this problem by adapting logical positivism's general approach to knowledge about

unobservables to the case of psychological states. The positivists held that inferences from observable facts about, say, falling bodies to unobservable facts about the gravitational field must be made via the use of analytically true "bridge laws," "coordinative definitions," or "meaning postulates" that give (at least part of) the meaning of "gravitational field" by listing the empirical consequences of various sentences in which the term appears. Behaviorism proposes something similar for desires (and beliefs). It is supposed to be *analytic* that any agent who desires one prospect more than another will be disposed to act in ways that are likely to produce the first rather than the second in any context where only one of the two can come about. This makes it a *misuse of language* to say that a person whose fair price for G is $\$g$ violates (a*)-(b*). The advantage of this rather radical stance is that it allows investigators to establish secure conclusions about a person's fair prices by watching what she does. When we see her giving up a sure fortune of $\$x$ to buy a wager in a situation where nothing else of value hangs in the balance, we can be certain that her fair price for the wager is greater than $\$x$, or so behaviorism has it.

Despite its attractiveness, the behaviorist interpretation of rational choice theory is slowly becoming a thing of the past (and should be made entirely a thing of the past as soon as possible). The old guard still insists that the concept of a fair price can only be understood in terms of behavioral dispositions, but it has become clear that the theoretical costs of this position far outweigh its benefits. There are just too many things worth saying that cannot be said within the confines of strict behaviorism. One cannot, for example, say whether an agent is disposed to satisfy (a*)-(b*) because she regards G and $\$g$ as equally desirable for their own sakes or because she believes that those who treat the less desirable G on a par with the more desirable $\$g$ will be rewarded in the hereafter. The basic difficulty here is that it is impossible to distinguish contexts in which an agent's behavior really does reveal what she wants from contexts in which it does not without appealing to additional facts about her mental state. Indeed, the only cases in which (a*)-(b*) are sure to indicate that a person finds G and $\$g$ equally desirable are those in which she is certain either that her choice will generate no consequences other than G or $\$g$ or that any consequences generated will have an equal effect on the desirabilities of these two propositions. To say this, however, is to go well beyond what the behaviorist interpretation allows since we are now imputing mental states to the agent that will not ultimately be manifested in her overt choice behavior.

An even more serious shortcoming is behaviorism's inability to make any sense of *rationalizing explanations* of choice behavior. When an agent is disposed to buy G for $\$x$ it always makes sense to inquire into her reasons. Does she have a *rationale* for making the purchase? If so, what is it? The answer, of course, is usually going to be that she believes this transaction will result in her gaining G and losing $\$x$, *and she finds this state of affairs preferable to one in which she does not gain G and keeps $\$x$* . If this explanation is to avoid circularity it must be possible to understand the italicized phrase as meaning something other than that she is willing to exchange G for $\$x$. So, insofar as we want to be in the business of giving rationalizing explanations of behavior, we must

recognize that there is a conceptual difference between finding G and $\$g$ equally desirable and being disposed to buy or sell G for $\$g$. More generally, since we shall often want to use (a)-(b) as a part of a noncircular rationalizing explanation for (a*)-(b*) we cannot simply regard the two as different ways of saying the same thing.

None of this is meant to deny that there is a tight connection between fair prices and betting behavior. An agent who has $\$g$ as her fair price for G will often be willing to buy the wager at prices not greater than $\$g$ and sell it at prices not less than $\$g$. Moreover, under the right conditions one *can* use this fact to draw reliable inferences about her fair prices on the basis of her behavior. However, unlike the behaviorist, who treated such inferences as analytic, we must recognize them as *inductions* from facts about choices to distinct facts about what the agent finds desirable. A person's betting behavior is often a highly reliable indicator of her fair prices, but it does not constitutively determine them.

The moral here is that rational choice theory is not so much about overt choices as about the underlying *desires and beliefs* that cause and rationalize them. When we criticize or commend any action on grounds of practical rationality we are not assessing the actor's behavior per se, but are evaluating the pattern of desires and beliefs that brought it about. More precisely, when we say that it was irrational, in the instrumental sense, for an agent to perform A we mean that she had an all-things-considered desire to do A , this desire was her motive for doing A , and it was irrational for her to want A given her other desires and beliefs. So, whenever some decision theorist says that a given action is irrational what he or she must really mean is that the agent's *all-things-considered desire to perform it* is irrational (in the sense that it does not cohere with her other beliefs and desires). It rarely does harm to speak of acts as rational and irrational without mentioning the underlying desires, but this is only because people so often do what they want. We must keep in mind that the connection between wanting and doing is not analytic. An inference from "The agent has an all-things-considered desire to perform A " to "She would perform A if given the chance" can be very strong in the inductive sense, as can its converse, but neither is grounded in any kind of conceptual necessity. For these reasons, our official version of Pascal's Thesis will be the one expressed in terms of desire rather than action.

1.4 AN ARGUMENT FOR PASCAL'S THESIS

Pascal never offered any systematic rationale for the policy of using expected payoffs as fair prices. This is not surprising. Proponents of new scientific ideas invest almost all their time in problem-solving activities; foundational issues only come to the fore after all the low-hanging fruit has been picked. With the hindsight of those fortunate enough to live after Frank Ramsey, Bruno de Finetti, Leonard Savage, John von Neumann, and Otto Morgenstern, it is now possible to see how a justification for Pascal's Thesis might go.⁹ We cannot, of course, be sure that it is the sort of justification Pascal would have offered had he thought seriously about the matter. (In fact, it is likely that he would have given up on the idea that fair prices are expected payoffs once he fully understood what it

entailed.) Be this as it may, the following argument does present the best case that can be made for Pascal’s Thesis. It is crucial that the reader have a clear sense of how it goes because a large part of this book will be spent improvising on its basic theme. I will present the argument rather pedantically, highlighting every relevant premise, so as to render the overall structure of the reasoning as transparent as possible.

Let’s begin by saying something about the set \mathbf{G} of wagers over which fair prices are to be defined. The argument requires \mathbf{G} to be a *mixture space*. An *even* or $(1/2, 1/2)$ mixture of two wagers G and G^* is a third wager H that always pays out a fortune of $\$x$ with probability $1/2\rho_1 + 1/2\rho_2$ when G_1 pays $\$x$ with probability ρ_1 and G_2 pays $\$x$ with probability ρ_2 . To illustrate, suppose that a nickel and a dime are going to be tossed, and consider

Nickel/Dime	Head/Head	Head/Tail	Tail/Head	Tail/Tail
$\rho =$	1/4	1/4	1/4	1/4
G_1	\$3	\$2	\$3	\$2
G_2	\$5	\$6	\$5	\$6
H_1	\$3	\$2	\$5	\$6
H_2	\$6	\$3	\$2	\$5

Here *both* H_1 and H_2 are even mixtures of G_1 and G_2 because each yields G_1 ’s payoffs with 1/2 probability and G_2 ’s payoffs with 1/2 probability. One can often think of such a mixture as a compound wager that offers G_1 and G_2 as “prizes” with equal probability. The third line of the table suggests such an interpretation: H_1 pays G_1 if the nickel falls heads and G_2 if it falls tails. This way of looking at things has limits, however, since there are mixtures, like H_2 , that cannot be understood as compounds.¹⁰

We can generalize the notion of an even mixture by imagining that the nickel in our example is not fair. If its bias toward heads is $\lambda/(1 - \lambda)$ where $1 \geq \lambda \geq 0$ (so that the ρ row of the table is: $\lambda/2, \lambda/2, (1 - \lambda)/2, (1 - \lambda)/2$), then H_1 , but not H_2 , is a $(\lambda, 1 - \lambda)$ mixture of G_1 and G_2 . Here is the formal definition:

H is a $(\lambda, 1 - \lambda)$ mixture of G_1 and G_2 if and only if, for every real number $x > 0$, H pays $\$x$ with probability $\lambda\rho_1 + (1 - \lambda)\rho_2$ whenever G_1 pays $\$x$ with probability ρ_1 and G_2 pays x with probability ρ_2 .

Mixtures have an illuminating geometrical interpretation. If we think of G_1 and G_2 as “points” in the “space” of all wagers \mathbf{G} , then a $(\lambda, 1 - \lambda)$ mixture of G_1 and G_2 lies on the “line segment” connecting them. An even mixture lies at the midpoint of the segment, a $(3/4, 1/4)$ mixture lies halfway between G_1 and the midpoint, a $(1/4, 3/4)$ mixture lies halfway between G_2 and the midpoint, and so on.

We can now state our first premise.

Premise 1 (Richness). (i) For any sum of money $\$x$, \mathcal{G} contains a wager that pays $\$x$ in every possible circumstance; (ii) \mathcal{G} is closed under mixing in the sense that for any $G_1, G_2 \in \mathcal{G}$ and $\lambda \in [0,1]$ there is $H \in \mathcal{G}$ such that, for every real number x , H pays $\$x$ with probability $\lambda\rho_1 + (1 - \lambda)\rho_2$ when G_1 pays $\$x$ with probability ρ_1 and G_2 pays $\$x$ with probability ρ_2 .

This says that the set of wagers is a *mixture space*: It contains all straight payoffs (think “constant wagers”), and it contains any mixture of any wagers it contains.

Pascal’s Thesis presupposes that each wager in \mathcal{G} will have a *single* fair price that holds good for all agents. There are two independent claims here.

Premise 2 (Existence). Each rational professional gambler will have a unique fair price for every wager in \mathcal{G} , that is, for each $G \in \mathcal{G}$ there must be a sum of money $\$g$ such that she finds holding G and having a fortune of $\$g$ equally desirable.

Premise 3 (Invariance). Each rational professional gambler must have *the same* fair price for *every* wager in \mathcal{G} .

Existence requires bettors to have highly *determinate* desires, so determinate, in fact, that they can put a precise price on every wager they consider. This leaves no room at all for vagueness or imprecision in judgments of value; for every sum of money $\$x$ there must be a fact of the matter about whether the bettor prefers G to $\$x$, prefers $\$x$ to G , or finds the two prospects equally desirable. Invariance forces every bettor to value money in the same way. One way to put this is to say that a fixed sum of money should “buy the same amount of happiness” for everyone.¹¹ Neither of these requirements is very plausible, and we shall reject both of them later.

The next premise is much more compelling. It simply says that rational professional bettors place a higher value on wagers that are sure to pay them more money.

Premise-4 (Dominance). If one wager $G = \&_S(S \square \rightarrow \$g_S)$ *dominates* another $H = \&_S(S \square \rightarrow \$h_S)$ in the sense that $g_S \geq h_S$ for all states S , then G ’s fair price is at least as great as H ’s. If, in addition, $g_S > h_S$ for all states S that entail some event E with positive probability, then G ’s fair price is strictly greater than H ’s.

Assuming that our gambler always prefers more money to less, this just says that she should never put a higher price on G^* than on G unless there is some chance that G^* will leave her richer than G will. For future reference note that Dominance entails that the fair price of every wager must fall somewhere between its highest and lowest payoffs.

By way of introduction to the next premise consider the wagers

Nickel/Dime	Head/Head	Head/Tail	Tail/Head	Tail/Tail
ρ	3/8	1/8	3/8	1/8
G	\$30	\$20	\$10	\$30
G^*	\$10	\$30	\$30	\$20

where the dime is biased 3:1 in favor of heads. Notice that G and G^* offer the same prizes with identical probabilities. As a consequence, someone who does not take them to be equally desirable must be basing her preference on a distinction between the two wagers that cannot be traced to differences in their payoffs or probabilities. She might, for example, prefer having a fortune of \$30 when the coins fall the same way to having that fortune when one comes up heads and the other comes up tails. Differences of this kind should only matter to a person whose interests in events outrun her interests in the fortunes they might produce. This never happens with professional gamblers. They treat money as an unalloyed good, so that for any event E and any wealth level $\$x$ they find the prospect ($\$x \& E$) exactly as desirable as the prospect ($\$x \& \neg E$). Our next premise codifies this point.

Premise 5 (Stochastic Equivalence). If G and H are *stochastically equivalent* in the sense that, for every real number x , G pays $\$x$ with probability ρ if and only if H also pays $\$x$ with probability ρ , then G and H have the same fair price.

The effect of this is to rule out differences in desirability among wagers that differ only in the identities of the chance events that underlie their probability assignments.

This makes it possible for us to think of fair prices as attaching not to individual elements of \mathbf{G} but to classes of stochastically equivalent wagers, where each such class may be identified with the probability distribution that is common to all its wagers. The main advantage of this is that it lets us pretend that a single object, $\lambda G_1 \oplus (1 - \lambda)G_2$, is *the* (λ , $1 - \lambda$)-mixture of G_1 and G_2 . $\lambda G_1 \oplus (1 - \lambda)G_2$ is thus a representative for any wager of the form

	E	$\neg E$
	λ	$1 - \lambda$
$\lambda G_1 \oplus (1 - \lambda)G_2$	G_1	G_2

What $\lambda G_1 \oplus (1 - \lambda)G_2$ really is, of course, is the probability distribution common to every wager H that can be expressed as a (λ , $1 - \lambda$)-mixture of any wagers stochastically equivalent to G_1 and G_2 . But, since the identities of the events in H do not matter except insofar as they have the probabilities λ and $1 - \lambda$, it makes sense to speak about these probabilities themselves, and thus to treat $\lambda G_1 \oplus (1 - \lambda)G_2$ as a single wager.

The following three principles express the basic formal properties of mixtures thought of in this abstract, event-independent way:

Mix₁. $G_1 = 1 \cdot G_1 \oplus 0 \cdot G_2$

Mix₂. $\lambda G_1 \oplus (1 - \lambda)G_2 = (1 - \lambda)G_2 \oplus \lambda G_1$

Mix₃. If $H = (\mu G_1 \oplus (1 - \mu)G_2)$, then $(\lambda H \oplus (1 - \lambda)G_2) = (\lambda\mu G_1 \oplus (1 - \lambda\mu)G_2)$.

All three are easily proven consequences of Stochastic Equivalence.

Our next premise is the linchpin of the entire theory, for it is what ensures that fair prices are governed by the laws of mathematical expectation. To motivate it, let's look at the following five wagers (where reference to the events have been suppressed because they do not matter):

ρ	3/8	1/8	3/8	1/8
G	\$100	\$40	\$100	\$40
G_1	\$30	\$20	\$30	\$20
G_2	\$10	\$50	\$10	\$50
H_1	\$30	\$20	\$100	\$40
H_2	\$10	\$50	\$100	\$40

Observe that $H_1 = 1/2G_1 \oplus 1/2G$ and $H_2 = 1/2G_2 \oplus 1/2G$. Since both wagers offer G as a “prize” with probability 1/2 it is natural to think that this common component should “drop out,” so that any difference in desirability between H_1 and H_2 would be traceable to differences in the desirabilities of G_1 and G_2 . This is what our next premise tells us.

Premise 6 (Independence). For $\lambda > 0$, the fair price of $H_2 = \lambda G_1 \oplus (1 - \lambda)G$ exceeds (is equal to) the fair price of $H_2 = \lambda G_2 \oplus (1 - \lambda)G$ if and only if the fair price of G_1 exceeds (is equal to) the fair price of G_2 .

This imposes a kind of noncontextuality constraint on rational desires. Since the identity of G is immaterial, the principle says that the relative strengths of a bettor's desires for G_1 and G_2 should not vary when they are mixed in the same proportion with any third wager. This makes it irrational to give more weight to the desirability of G_1 than to that of G_2 when the two appear in $H_1 = \lambda G_1 \oplus (1 - \lambda)G$ and $H_2 = \lambda G_2 \oplus (1 - \lambda)G$ but to weigh G_2 more heavily than G_1 when they appear in, say, $H_1^* = \lambda G_1 \oplus (1 - \lambda)G^*$ and $H_2^* = \lambda G_2 \oplus (1 - \lambda)G^*$. We will discuss some of the reasons for accepting this principle in the next chapter.

In the presence of the other premises Independence entails the following important consequences (whose proofs are straightforward and will be left to the reader¹²):

Lemma 1.1a (Substitution). $H_1 = \lambda G \oplus (1 - \lambda)G^*$ and $H_2 = \lambda g \oplus (1 - \lambda)G^*$ have the same fair price when g is G 's fair price.

Lemma 1.1b (Stochastic Dominance). If G 's fair price is greater than G^* 's, then the fair price of $H = \lambda G \oplus (1 - \lambda)G^*$ increases monotonically with λ .

Lemma 1.1C (Averaging). If $\lambda, \mu, \nu \in [0,1]$ and if

- g_λ is the fair price of $G_\lambda = \lambda x \oplus (1 - \lambda)y$
- g_μ is the fair price of $G_\mu = \mu x \oplus (1 - \mu)y$
- h is the fair price of $H = \nu g_\lambda \oplus (1 - \nu)g_\mu$

then h is also the fair price of $H^* = [\sigma x \oplus (1 - \sigma)y]$ where $\sigma = (\nu\lambda + (1 - \nu)\mu)$.

According to Lemma 1.1a, it should never matter to a rational gambler whether she stands to get G with a certain probability or its fair price with that same probability. 1.1b is a formal way of saying that prospects offering higher probabilities of better outcomes are more desirable. While 1.1c looks complicated, all it really says is that the fair price of a $\nu/(1 - \nu)$ mixture of the fair prices of G_λ and G_μ is also the fair price of a $\nu/(1 - \nu)$ mixture of G_λ and G_μ themselves.

As a consequence of Lemma 1.1b, the fair price of $G_\rho = \rho x \oplus (1 - \rho)y$ increases monotonically in ρ when $x > y$. Our next premise ensures that this increase is a continuous one, so that small changes in probability never lead to large changes in G_ρ 's fair price.

Premise 7 (Continuity). Given any real numbers $x_1 > y > x_0$, there exists a probability ρ_y such that y is the fair price of $G = \rho_y x_1 \oplus (1 - \rho_y)x_0$.

Lemma 1.1b ensures that ρ_y is unique.

We are now in a position to state what is perhaps the single most important result in expected utility theory:

Theorem 1.2. Given any real scaling constants $u > z$, if Premises 1-7 hold then there exists a unique monotonically increasing real function u such that

- $u(u) = 1$ and $u(z) = 0$
- The fair price g for any wager $G = \&_S(S \mapsto g_S)$ is the unique solution to the equation $u(g) = U(G)$ where $U(G)$ is the *expected value* of $u(g_S)$ computed relative to ρ , so that $U(G) = \int_S u(g_S)d\rho$, or, when the set of states is countable, $U(G) = \sum_S u(g_S)\rho(S)$.

The proof of this result is too technical to be conveniently included in an introductory chapter of this sort. Readers interested in the details may consult any one of a number of

excellent sources,¹³ or they might just read on since subsequent chapters will present proofs of more general theorems that entail this one as a special case.

The map u is not something Pascal would have recognized. It is called a *utility* function for money and should be interpreted as a measure of the overall “amount” of happiness or satisfaction that various sums of money would be able to buy (for a professional bettor). Theorem 1.2 shows that, subject to the arbitrary choice of a zero point z and unit u to set the scale for measuring utility, Premises 1-7 guarantee the existence of a unique function u that makes the expected utility of any wager coincide with the utility of its fair price. This means that a bettor who satisfies the premises will set fair prices as if she is calculating each wager G 's expected utility and then aiming to find a sum of money $\$g$ whose utility is the same as $U(G)$. This is not to say that she actually employs this highly abstract process when she goes about fixing fair prices. The point is only that from the *third person perspective* it will look *as if* her fair prices have been arrived at in this way.

One illuminating way to understand Theorem 1.2 is by recognizing that the function u provides an *interval scale* on which to measure the “value” of money.¹⁴ To see what this involves, let G be a wager that offers fortunes of $\$x$ and $\$y \neq \x with probabilities ρ and $1 - \rho$ respectively. For two utility functions u and u^* to generate the same fair price for G there would need to be a real number $x \leq g \leq y$, such that $u(\$g) = \rho u(\$x) + (1 - \rho)u(\$y)$ and $u^*(\$g) = \rho u^*(\$x) + (1 - \rho)u^*(\$y)$. This can only happen if

$$\frac{u(\$g) - u(\$y)}{u(\$x) - u(\$y)} = \rho = \frac{u^*(\$g) - u^*(\$y)}{u^*(\$x) - u^*(\$y)}$$

In other words, the *ratio* of the *difference* in utility between $\$g$ and $\$y$ to the *difference* in utility between $\$x$ and $\$y$ must be an *invariant* quantity among all functions that satisfy $u(g) = U(G)$. The reader is invited use this observation to verify that the following three conditions are equivalent.

(A) For all wagers $G \in \mathbf{G}$ and all real numbers x , $u(x) \geq U(G)$ if and only if $u^*(x) \geq U^*(G)$.

(B) Ratios of expected utilities are invariant between u and u^* , that is, for any wagers G_1, G_2, G_3 and G_4 with $\$g_3 \neq \g_4 one has

$$\frac{U(G_1) - U(G_2)}{U(G_3) - U(G_4)} = \frac{U^*(G_1) - U^*(G_2)}{U^*(G_3) - U^*(G_4)}$$

(C) u and u^* are positive linear transformations of one another, so that $u^*(x) = m u(x) + b$ where $m > 0$ and b are constants. These constants can be written as $m = [u^*(x_1) - u^*(x_2)]/[u(x_1) - u(x_2)]$ and $b = [u^*(x_1)u(x_2) - u(x_1)u^*(x_2)]/[u(x_1) - u(x_2)]$ for any real numbers x_1 and x_2 with $u(x_1) > u(x_2)$.

(A)-(C) tell us that any utility function that yields the right fair prices for wagers *cardinally measures* the value of money. This means, first, that its expected values are *ordinally significant*, so that $U(G) > U(H)$ holds only if all rational professional bettors really do prefer holding G to holding H . Second, *ratios of differences* of these expected values have *cardinal significance*; they accurately express the amount by which various changes in a bettor's portfolio of wagers would contribute to her happiness. One way to put this is to say that the identity $U(G_1) - U(G_2) = \lambda[U(G_3) - U(G_4)]$ holds only when the change in happiness caused by gaining G_1 and losing G_2 is λ times the change caused by gaining G_3 and losing G_4 . Last, ratios of u 's expected values lack cardinal significance, so that $u(G) = 2u(G^*)$ cannot be taken to mean that G is twice as desirable as G^* . In light of all this, we may rewrite Theorem 1.2 as follows: *If a professional gambler's fair prices satisfy Premises 1-7, then there will be an interval scale, unique up to positive linear transformation, that quantifies the "value" of money for her, and she will price wagers according to their expected "values" as measured on this scale.*

We will discuss the concept of utility more fully in the next section, but we first must understand why it did not enter into Pascal's thinking. Pascal implicitly treated differences in the sizes of fortunes as differences in the amount of happiness they can buy. So, while the letter of his position commits him to all the premises used in our argument thus far, his conclusion requires one more to ensure that $u(x) = x$. The further premise simply says that Pascal's Thesis holds in the special case of an even odds bet.

Premise 8 (Risk Neutrality). $\$(x + y)/2$ is the fair price for a wager that pays $\$x$ or $\$y$ with equal probability.

In the presence of the other premises this turns out to be enough to ensure that fair prices go by expected payoffs generally.

Premise 8 can be given a plausible sounding rationale. Consider three wagers of the form

	ρ	$1 - \rho$
G_0	$\$x$	$\$y$
G_1	$\$(x + y)/2$	$\$(x + y)/2$
G_2	$\$y$	$\$x$

Cover up the G_2 row and observe that the difference between the payoffs of G_0 and G_1 in the first column is $(x - y)/2$, and in the second is its negative $(y - x)/2$. Now cover the G_0

row and note that the difference between the payoffs of G_1 and G_2 in the first column is $(x - y)/2$, and in the second is $(y - x)/2$. Thus, the change in the agent's fortune in moving from G_0 and G_1 is the same as the change in her fortune in moving from G_1 and G_2 under both the ρ -event and the $(1 - \rho)$ -event. Given this symmetry, it is reasonable to think that the difference in value between G_0 and G_1 should match the difference in value between G_1 and G_2 . Accordingly, the fair prices should line up so that $\$g_1 = \$(x + y)/2$ falls halfway between $\$g_0$ and $\$g_2$. Now, in the case where $\rho = 1/2$ we know that $\$g_0$ and $\$g_2$ will be identical (by Stochastic Equivalence), so $\$g_0 = \$g_2 = \$g_1 = \$(x + y)/2$ as Premise 8 requires.

The fallacy here lies in the step from “the change in the agent's *fortune* in moving from G_0 and G_1 is the same as the change in her *fortune* in moving from G_1 and G_2 ” to “the difference in *value* between G_0 and G_1 should match the difference in *value* between G_1 and G_2 .” This inference only goes through if we assume that

Money has Constant Marginal Utility. For any real numbers $x > y \geq 0$, the difference in desirability to a rational professional gambler between having a fortune of $\$y$ and having one of $\$(x + y)/2$ should be the same as the difference in desirability between having a fortune of $\$(x + y)/2$ and having one of $\$x$.

Among other things, this entails that a rational gambler should always desire a fortune of $\$2f$ twice as strongly as a fortune of $\$f$, and that her desire for an extra dollar should never wax or wane with changes in her fortune, so that the difference in desirability between $\$f$ and $\$(f + 1)$ is the same for every f . If this does not seem any more obvious to you than Premise 8, do not be alarmed – the two principles are equivalent in the context of the other premises, and neither is true.

1.5 THE CRAMER/BERNOULLI THESIS

Gabriel Cramer and Daniel Bernoulli deserve the credit for discovering that money does not have constant marginal utility (though Cramer's claim is often ignored).¹⁵ It was Cramer who first realized that fair prices cannot plausibly be identified with expected payoffs. He came to this conclusion during the course of investigations into a gambling problem proposed by Nicholas Bernoulli's brother, Daniel, which came to be called the *St. Petersburg Paradox*. Suppose a fair coin is going to be tossed until a tail comes up, and let E_k be the event of the first tail coming up on the k th flip. Consider a wager of the form $G_f = \&_k(E_k \rightarrow \$(f + 2^k))$, for $k = 1, 2, 3, \dots$ and $f > 0$, where $\$f$ is your present net worth. If you hold G_f your net worth will increase by $\$2$ if the first tail comes up on the first toss, by $\$4$ if the first tail comes on the second toss, of $\$8$ if the first tail comes on the third toss, and so on. What do you think G_f 's fair price ought to be? That is, at what value $g_f = (f + \delta_f)$ would you be equally happy having $\$g_f$ as your net worth or letting your fortunes ride on G_f ? When asked this question most people set $\$g_f$ at roughly $\$20$ above their current fortune. That is, they profess to be indifferent between having an extra $\$20$ in their pocket or holding a bet that offers them an extra $\$2$ with probability

1/2, an extra \$4 with probability 1/4, an extra \$8 with probability 1/8, and so on. There are variations, of course, but nearly everyone chooses a low number, far lower than, say, \$100,000. This is not surprising since at $g_f = (f + 100,000)$ the chance of G_f paying more than g_f is about 0.0000076, and the chance of it paying at least \$90,000 *less* than g_f is 0.9998.

The trouble is that a person who identifies g_f with $\text{Exp}_{\$}(G_f)$ should think it wise to forgo *any* finite, risk-free increase in wealth payment to obtain G_f because G_f has an *infinite* expected payoff: $\sum_k \rho(E_k)(f + 2^k) = f + \sum_k (1/2^k)2^k = \infty$. She should, for example, prefer playing the St. Petersburg game to having a straight payment of a *trillion* dollars even though she is fully aware that the coin would need to come up heads *forty* times in a row for her to win that much. (To put this in perspective, the odds of tossing forty straight heads with a fair coin have about the same order of magnitude as the odds of a large asteroid hitting the earth in the next second.) Pascal's Thesis thus leads to patently absurd results when applied to the St. Petersburg game.¹⁶

The root of the problem, as Cramer recognized, is that the identification of fair prices with expected payoffs only makes sense if money has constant marginal utility. In the St. Petersburg game Pascal's Thesis requires a rational bettor to regard a potential increase of $\$2^{k+1}$ in her fortune as being twice as desirable as an increase of $\$2^k$. If she did not, there would be no reason to think that the former prospect should have the same effect on G_f 's value when discounted by a probability half as large. This is just a special case of the general claim, inherent in the idea that money has constant marginal utility, that the difference in desirability between having a fortune of $\$x$ (read $\$f$) and having one of $\$(x + y)/2$ (read $\$(f + 2^k)$) should be the same as the difference in desirability between having a fortune of $\$(x + y)/2$ and having one of $\$y$ (read $\$(f + 2^{k+1})$). This was Pascal's mistake.

Money, contrary to the old adage, can buy happiness¹⁷ (even if there are many things that contribute to happiness that it cannot buy), but the "rate of exchange" between the two is not constant. The "happiness value" of a given dollar depends on how many others one will have left once it is spent; money is always worth more to a pauper than to a prince. By identifying the fair prices with expected payoff Pascal was implicitly denying this. On his view the value of an extra dollar should be invariant among all professional gamblers *no matter how large or small their fortunes happen to be*. Cramer recognized this as a mistake; the "amount of happiness" that $\$2^{k+1}$ can buy is usually less than twice the amount $\$2^k$ can buy, far less when k gets large. In effect, then, Cramer rejected the idea that money has constant marginal utility, and with it Premise 8 of our argument for Pascal's Thesis.

Once one does this a way of resolving the St. Petersburg Paradox suggested itself almost immediately. If the "value" of having $\$2^k$ rises slowly enough as k grows, then there can be a finite amount of money that would buy the same "amount" of happiness as G_f can be expected to buy. This would make g_f finite even though G_f 's expected payoff is infinite. Of course, for this solution to work there needs to be a way of making sense of talk about

the “value” of money or the “amount” of happiness it buys. Thus it was that Cramer first hit upon the idea of a real function $u(x)$ that would measure the overall desirability of having a fortune of $\$x$. Let’s call Cramer’s function a “utility” even though this term was not used until long after his death. (Also, for historical accuracy, it should be noted that Cramer saw u as measuring the value of *changes* in a bettor’s net worth rather than the value of her total fortune.)

Once one starts thinking in terms of utility rather than money it is a short step to the view that each wager should be valued by its *expected utility* $U(G) = \int_S u(g_S)d\rho$ (or $U(G) = \sum_S u(g_S)\rho(S)$ when the set of states is countable) rather than its expected payoff. Cramer thus took $U(G)$ to be the right measure of the “amount of happiness” that a gambler can reasonably expect to gain from holding G . This led him to replace Pascal’s Thesis by

The Cramer/Bernoulli Thesis. If the utility $u(x)$ accurately measures the amount of happiness or satisfaction that a fortune of $\$x$ can buy, if $U(G)$ is the expected utility of the wager G (computed relative to ρ), and if $\$g$ is a sum of money such that $u(g) = U(G)$, then any professional gambler should find the prospect of holding G and the prospect of having a fortune of $\$g$ equally desirable.

If this is right, then the St. Petersburg Paradox can be solved simply by showing that the correct utility function for money makes $\sum_k 1/2^k u(2^k)$ finite even when $\sum_k (1/2^k)2^k$ is infinite. Cramer knew that $u(2^{k+1}) < 2u(2^k)$ was a sufficient condition for $U(G_f) < \infty$,¹⁸ and he maintained that any reasonable utility should obey the former inequality. He did not go on to defend any specific u as *the* correct utility function, though he did point out that the expected utility of the St. Petersburg game would be finite when $u(\$x) = x^{1/2}$ or when $u(2^k)$ is constant for all $k > 24$.

By requiring that $u(2^{k+1}) < 2u(2^k)$ Cramer was, in effect, appealing to the fact that

Money has Declining Marginal Utility.¹⁹ For any real numbers $x > y$, the difference in desirability, from a professional gambler’s point of view, between having a fortune of $\$x$ and having one of $\$(x + y)/2$ should be greater than the difference in desirability between having a fortune of $\$(x + y)/2$ and having one of $\$y$.

to provide a rationale for replacing Premise 8 of the previous section by

Premise 8* (Risk Aversion). The fair price for a wager that pays $\$x$ and $\$y$ with equal probability should be strictly less than $\$(x + y)/2$.

Given the other premises, Premise 8* implies that the gambler should be *risk averse* at all wealth levels. In the jargon of modern economics, a person is *risk averse*, *risk neutral*, or *risk seeking* at wealth level $\$f$ just in case there is positive number ϵ such that for any $0 <$

$\delta < \varepsilon$ she would find having a sure fortune of $\$f$ strictly more, exactly as, or strictly less desirable than having the wager

$$G(f, \delta) \quad \begin{array}{c} 1/2 \\ \$(f + \delta) \end{array} \quad \begin{array}{c} 1/2 \\ \$(f - \delta) \end{array}$$

A risk averter will set a price $\$g(f, \delta)$ for $G(f, \delta)$ that is less than its expected payoff $\$f$, a risk neutralist will set $\$g(f, \delta) = \f , and a risk seeker will set $\$g(f, \delta) > \f . Pascal's Thesis gives absurd results because it assumes that all rational bettors are risk neutral everywhere. Cramer's great insight was to see that the hypothesis of universal risk neutrality should be replaced by the hypothesis of universal risk aversion.

Daniel Bernoulli gets part of the credit for the Cramer/Bernoulli Thesis because he was first to suggest a plausible candidate for u , first to appreciate that utilities attach to wealth levels rather than changes in wealth, and first into print. Bernoulli, who ranks among the foremost mathematicians of the eighteenth century (even though he was only mediocre for a Bernoulli), learned of the St. Petersburg problem by reading a letter from Cramer to his *cousin* Nicholas Bernoulli²⁰ outlining the expected utility hypothesis. Daniel immediately recognized the merit in Cramer's proposal and went beyond it by arguing that utility must be a *logarithmic* function of money. He arrived at this conclusion by supposing that the utility gained by increasing a fortune of $\$f$ by an increment of $\$\Delta f$ should always be inversely proportional to the fortune's size and directly proportional to that of the increment, so that $u(f + \Delta f) - u(f) = \alpha \Delta f / f$ for some positive α . The only functions with this property have the logarithmic form $u(x) = \alpha \log(x) + \beta$, where $\alpha > 0$ and β are real constants. These provide a system of interval scales for measuring the value of money with different choices of α and β indicating different conventions for setting the unit and zero. As was noted in the previous section all such scales determine the same fair prices, so we lose nothing by simply setting $u(x) = \log(x)$.

This function is risk averse everywhere,²¹ and it produces finite fair prices for the St. Petersburg game that seem fairly reasonable. Here is a sampling:

Value of f	Fair price for G_f	Amount by which $\$G_f$ exceeds $\$f$
\$1,000	\$1012	\$12.00
\$2,000	\$2,013	\$13.00
\$10,000	\$10,015.25	\$15.25
\$20,000	\$20,016.25	\$16.25
\$100,000	\$100,018.50	\$18.50
\$200,000	\$200,019.50	\$19.50
\$1,000,000	\$1,000,021.50	\$21.50
\$2,000,000	\$2,000,022.50	\$22.50

Notice that the fair price of G_f increases with f . This corresponds to the often observed fact that the *degree* of risk aversion tends to decline with increases in an agent's fortune. (The rich, in other words, tend to be more willing to risk wealth than the poor in speculative ventures.)

While modern economists follow Cramer and Bernoulli in thinking both that money has declining marginal utility and that wagers should be valued by their expected utilities, they reject Bernoulli's claim that $u(x) = \log(x)$. The fact is that Bernoulli's definition does not really solve the St. Petersburg Paradox, but merely sweeps it under the rug. To see why, consider the wager $H_f = \&_j (E_j \rightarrow (f + \text{\$exp}(2^j)))$, where $\text{exp}(x)$ is the base of the natural logarithm raised to the x th power. On Bernoulli's proposal H_f has infinite expected utility. The problem with this is that anyone who assigns H_f infinite expected utility will also assign infinite utility to any arrangement $H_\rho = \rho H_f \oplus (1 - \rho)\text{\$}x$ in which she stands to gain H_f with some probability $\rho > 0$ (no matter how small) and a finite "consolation prize" of $\text{\$}x$ with probability $1 - \rho$. This causes a number of formal problems. For instance, Stochastic Dominance is violated because, even though it seems that a bettor holding H_ρ ought to do better as ρ gets larger, this is not borne out by the expected utilities, which are all infinite. Second, a wager of infinite utility will be strictly preferred to *any* of its payoffs since the latter are all finite. This is absurd given that we are confining our attention to bettors who value wagers only as means to the end of increasing their fortune.

All this aside, the real problem is that the infinite utility of H_f forces a decision maker completely to ignore facts about what might happen if she does not obtain H_f , even though these facts clearly should be relevant to her deliberations. No matter how close ρ is to zero, an expected utility maximizer will be unable to see any difference between a case in which the consolation prize is abject poverty ($x = 0$) and one in which it is fabulous wealth ($x = \text{\$}10,000,000$) since both *total* situations will seem infinitely desirable to her. Thus, an agent who has a chance at some infinitely desirable outcome does not need to consult her beliefs or other desires in coming to a decision about what to do or want. Her task is clear and simple: She should try to obtain the infinitely desirable thing by any means that affords her some chance of getting it, and *any one of these means is just as good as any other*. She must, in short, become blind to "merely finite" things in her pursuit of the infinite good. This is clearly irrational when a "merely finite" thing is all one is likely to get.

To see the point most vividly, consider Pascal's famous "wager" argument, which tries to use expected utility theory, in a qualitative way, to convince people to go to church, take the sacraments, and lead upright lives. The claim is that doing these things is likely to lead to belief in God, which increases one's chances of securing the infinite rewards of paradise. There are many well-known criticisms of Pascal's reasoning, but the most telling is one due to Anthony Duff and Alan Hajek.²² While neither Duff nor Hajek puts things quite this way, the basic point is that Pascal's conclusion does not follow from his

premises because even if going to church, taking the sacraments, and living an upright life did give one a chance at infinite happiness, this would not be a reason to prefer such a life to one that involves torturing and killing those who go to church, take sacraments, and live uprightly because, as the case of St. Paul amply demonstrates, there is a nonzero probability that this too will secure one a place in the New Jerusalem. It would do no good for Pascal to respond that the probability of eternal bliss is higher for do-gooders than for those who persecute them. After all, by his own lights, the magnitudes of these probabilities become irrelevant when they are multiplied by an infinite utility. This, indeed, is the very point on which his argument hinges.

There are three ways to prevent wagers from having infinite fair prices so as to avoid any version of the St. Petersburg Paradox. The simplest is to claim that the expected utility analysis of fair prices only makes sense when it is applied to wagers based on *finite* event partitions. This solves the problem, but only at the expense of imposing a limitation on decision theory that seems far too extreme. It is perfectly possible, after all, for us to make sense of wagers that have infinitely many possible outcomes. As a second option, one might allow unbounded utility functions, but restrict the allowable gambles in such a way that events of high utility are always assigned such low probabilities that infinite expected utilities never arise.²³ The St. Petersburg Paradox would not be a problem on this view because there would be no chance that anyone would ever be faced with it. A third approach, the one I prefer, is to require that a rational agent's utility function for money be *bounded* in the sense that there is a $b > 0$ such that $-b < u(\$x) < b$ for all $x \in [0, \infty]$. Under these conditions every wager the agent considers will have a finite expected utility, and the sorts of problems we have been discussing will not arise. Since $\alpha \log(x) + \beta$ fails all three of these tests it cannot be the correct utility function for money (though it may be a good approximation in regions far from infinity). This is one reason for rejecting Bernoulli's definition of utility. There is another.

1.6 INSTRUMENTAL RATIONALITY AS SUBJECTIVE EXPECTED UTILITY MAXIMIZATION

A second reason for rejecting Bernoulli's definition is that it is wrong to think that any *one* definition of utility should hold good for everyone. There is no *single* right way for a professional gambler to value money. The desire for money is (to some extent) a matter of taste, and some people have more of a taste for it than others. Since rational choice theory cannot be in the business of dictating tastes, it should start by taking a gambler's desire for money as a datum, and go on to dictate *her* fair prices in terms of it. From the perspective of the theory there is no such thing as *the* correct utility function for money, *the* right fair price of a wager, or *the* right risk premium to pay at any level of wealth. A professional gambler's desire for money is ultimately an instance of *de gustibus non disputandum*.

Even though Bernoulli was wrong to insist that each rational gambler must have exactly the same utility function, he was right to think that there should always be some utility function or other that accurately measures the "amount of happiness" that fortunes of

various sizes will buy *for her*. While this would make it incorrect to say that a gamble $G = \sum_S (S \rightarrow \$g_S)$ has *the* fair price given by $\log(\$g) = \sum_S \rho(S) \log(\$g_S)$, it would remain true that for each rational bettor there must be some number that is G 's fair price *for her*. So, in backing away from Bernoulli's definition, one might still want to accept the Cramer/Bernoulli Thesis subject to the proviso that the appropriate utility function, and the fair prices associated with it, will vary from agent to agent.

This leaves us with the problem of saying which utilities go with which agents. Let us think of each gambler as being characterized by a set of statements that describe the strengths of her desires for wealth, and wagers involving wealth. We will call these statements *constraints*²⁴ because they constrain the form of the utility function that is used to measure the strength of the agent's desires. The most common constraints, and the ones that will be most important in the present context, specify the strengths of the gambler's desires in *comparative* terms by saying things like the following:

- She strictly prefers holding G to holding G^* .
- Her fair price for G is $\$g$; that is, she is indifferent between holding G and having a guaranteed fortune of $\$g$.
- She is everywhere risk averse.

Other constraints have a *cardinal* character. They might say things like the following:

- The difference between the strength of the gambler's desire for G and the strength of her desire for G^* is twice the difference between the strength of her desire for H and the strength of her desire for H^* .

A utility u will be said to *represent* the bettor's desires if anyone who formed desires about wagers by maximizing expected u values would automatically satisfy all the constraints that the bettor's desires satisfy. Thus to satisfy the constraints listed above we would need a utility u for which

- $U(G) > U(G^*)$
- $u(\$x) = U(G)$
- $u(\$f) > U(G(f, \delta))$ for all f and small δ
- $U(G) - U(G^*) = 2[U(H) - U(H^*)]$

Once one rejects the notion of a one-size-fits-all utility function the next natural step is to require that every rational agent's fair prices be described by a utility that represents her desires.

Decision theorists have typically been concerned with ordinally specified constraints on desire, what are often called *preferences*. We think of these as an agent's "all-things-

considered” judgments about the desirability of wagers in \mathbf{G} . Since \mathbf{G} contains all the *constant* wagers that pay out the same fixed sum $\$x$ under every contingency, the desirability of riskless as well as risky prospects can be modeled as preferences for wagers in \mathbf{G} . A bettor might be in any of the following six attitudinal states vis-à-vis the comparative desirability any pair of wagers G and G^* :

- G is *strictly preferred* to G^* (written $G \succ G^*$); from the agent’s perspective it would be more desirable to hold G than to hold G^* .
- G is *weakly preferred* to G^* ($G \succeq G^*$); from the agent’s perspective it would be at least as desirable to hold G as to hold G^* .
- The agent is *indifferent* between G and G^* ($G \approx G^*$); the agent finds the prospect of holding G or holding G^* equally desirable.
- G^* is *strictly preferred* to G ($G^* \succ G$).
- G^* is *weakly preferred* to G ($G^* \succeq G$).
- *No preference*: The strengths of the agent’s desires regarding G and G^* are not sufficiently determinate to warrant any of the previous characterizations.

The totality of these attitudes is the agent’s *preference ranking* over wagers.²⁵ Formally, this is a pair of binary relations (\succ, \succeq) defined on \mathbf{G} in which \succ encodes the agent’s strict preferences and \succeq encodes her weak ones. Indifference can be introduced via the definition: $G \approx G^*$ if and only if $G^* \succeq G$ and $G \succeq G^*$. The agent’s fair prices show up within (\succ, \succeq) as indifferences of the form $\$g \approx G$ where $\$g$ is thought of as a constant wager that pays $\$g$ in each state of the world.

If we think of each agent’s desires as being fully described by her preference ranking, then one natural way to apply the expected utility hypothesis is by making it a requirement of instrumental rationality that there should always be an agent-specific *subjective utility function* u defined on \mathbf{G} that *strongly ordinally represents* her preferences in the sense that, for any two wagers G and G^* in \mathbf{G} , the agent strictly (weakly) prefers G to G^* if and only if G ’s expected utility is greater than (not less than) G^* ’s utility where both these utilities are computed using u . Formally, the requirement would be as follows:

Subjective Expected Utility Hypothesis (strong version). In order for a “professional” gambler to be rational in the instrumental sense there must be at least one utility function for money, u , whose associated expectation operator, U , strongly ordinally represents the gambler’s preference ranking in the sense that both

$$G \succ G^* \text{ if and only if } U(G) > U(G^*) \text{ and}$$

$$G \succeq G^* \text{ if and only if } U(G) \geq U(G^*)$$

hold for any $G, G^* \in \mathbf{G}$.

On the reasonable assumption that an instrumentally rational bettor will always perform her most preferred option, this principle entails that such a bettor will invariably seek to obtain a wager that maximizes her subjective expected utility. Moreover, since \mathbf{G} contains all “constant” wagers, and since each $G \in \mathbf{G}$ can be associated with a unique sum g at which the bettor is indifferent between g and G , the identity $u(\$x) = U(G)$ will hold just in case $x = g$. Thus, the fair price for any wager will conform to its expected utility, just as Pascal, Cramer, and Bernoulli suggested, but it is now the bettor’s *subjective* expected utility that matters.

The classic defense of The Subjective Expected Utility Hypothesis was presented by John von Neumann and Otto Morgenstern in their immensely influential *Theory of Games and Economic Behavior*.²⁶ The argument they give is a generalization of Theorem 1.2 presented earlier. Von Neumann and Morgenstern set down a set of axioms on rational preference rankings whose satisfaction was sufficient to ensure that the ranking could be represented by an expected utility function. Here is a system of axioms that is equivalent to the set they presented:²⁷

VN₁ (\succ, \succeq) partially orders \mathbf{G} .

Reflexivity of Weak Preference. $G \succeq G$.

Consistency. $G \succ G^*$ only if $G \succeq G^*$
 $G \succeq G^*$ only if not $G^* \succ G$

Transitivity. If $G \succeq G^*$ and $G^* \succeq G^{**}$, then $G \succeq G^{**}$
 If $G \succ G^*$ and $G^* \succeq G^{**}$, then $G \succ G^{**}$
 If $G \succeq G^*$ and $G^* \succ G^{**}$, then $G \succ G^{**}$

VN₂ *Averaging.* If G ’s lowest payoff is at least as high as G^* ’s highest payoff, then $G \succeq G^*$. If, in addition, there is a nonzero probability that G and G^* have different payoffs, then $G \succ G^*$.

VN₂ *Independence.* $G \succ G^*$ if and only if $G \oplus G^{**} \succ G^* \oplus G^{**}$
 $G \succeq G^*$ if and only if $G \oplus G^{**} \succeq G^* \oplus G^{**}$

VN₃ *Stochastic Dominance.* Let G pay x_1 with probability ρ and $x_2 < x_1$ with probability $1 - \rho$ and let G^* pay x_1 with probability ρ^* and x_2 with probability $1 - \rho^*$. Then, $G \succeq G^*$ iff $\rho \geq \rho^*$.

VN₄ *Archimedes' Axiom.* If $G \succ G^* \succ G^{**}$, then there is a real number λ strictly between zero and one such that $\lambda G \oplus (1 - \lambda)G^{**} \succ G^* \succ (1 - \lambda)G \oplus \lambda G^{**}$.

VN₅ *Linearity:* For any real number $\lambda > 0$, $G \succeq G^*$ if and only if $\lambda G \succeq \lambda G^*$, and $G \succ G^*$ if and only if $\lambda G \succ \lambda G^*$.

VN₆ *Completeness:* Either $G \succ G^*$ or $G^* \succeq G$.

Von Neumann and Morgenstern showed how these axioms can be used to construct an expected utility representation for any preference ranking that satisfies them. Here is their result:

Theorem 1.3 (von Neumann/Morgenstern). If a preference ranking (\succ, \succeq) satisfies VN₁-VN₆, then there exists a utility function u for money whose associated expectation operator $U(G) = \sum_S \rho(S)u(g_S)$ strongly ordinally represents (\succ, \succeq) in the sense that

$$\begin{aligned} G \succ G^* &\text{ if and only if } U(G) > U(G^*) \text{ and} \\ G \succeq G^* &\text{ if and only if } U(G) \geq U(G^*). \end{aligned}$$

for any wagers $G, G^* \in \mathcal{G}$. Moreover, u is unique up to the arbitrary choice of a unit and a zero point relative to which utility is measured.

Theorem 1.3 tells us that a preference ranking defined over \mathcal{G} can only satisfy the von Neumann/Morgenstern axioms if it can be represented by a *cardinal* utility function whose expected values measure the “amount of overall happiness” that the wagers in \mathcal{G} can buy. Thus, anyone who holds that all of the von Neumann/Morgenstern axioms are strict requirements of instrumental rationality is thereby committed to the view that the strengths of a rational bettor’s desires can be measured on an interval scale with the same degree of precision that the temperatures of physical objects can be measured on the Celsius scale.

This is a strong requirement. Indeed, most proponents of expected utility maximization now think it is too strong. It commits what Mark Kaplan has called “the sin of false precision”²⁸ by supposing that the desires of any rational gambler will be sufficiently precise to be measured by real numbers. Since her preference ranking must be represented by a *single* utility function (once a unit and a zero are in place), there is no room for any kind of vagueness or indeterminacy in the strengths of her desires. Just as every physical object is taken to have a definite temperature, so every desire must have a definite strength. This is clearly unrealistic. Many desires lack determinate strengths, and this is no indication of the irrationality of those who hold them. Nonmonetary cases provide the most compelling examples: If hearing no music is 0 and hearing Mozart’s

Jupiter Symphony played by a first-class orchestra is 1, how much would it be worth, in terms of your overall happiness, to hear John Coltrane's *A Love Supreme* played by an excellent jazz quartet? Your tastes will probably dictate whether the answer should be "more than 1," "less than 0," "somewhere between 0 and 1," "closer to 1 than 0," or even, "very close to 1." I doubt, however, that you can come up with a number that gauges the strength of your desire to the third decimal place. This is not because you don't know what the number is; it is because there is no such number to be known. Even in the case of wagers with monetary outcomes it seems unlikely that our desires are anywhere near as definite as the von Neumann/Morgenstern axioms demand. Exactly how much is a fortune of \$100,000 worth if \$0 has utility 0 and \$200,000 has utility 1? Or, at precisely what value of ρ would you be indifferent between having a sure \$10,000 and a wager that pays \$20,000 with probability ρ and \$0 with probability $1 - \rho$? It is no sin against rationality not to know how to answer these questions, or for them not to have any answers at all.²⁹ The general point here is that once we give up on finding a one-size-fits-all utility we open up space for the idea that people can differ not only in what they want and how strongly they want it, but in the extent to which these wants are determinate.

Indeterminacy in a bettor's desires shows up as *incompleteness* in her preferences; the trichotomy property – $G \succ G^*$, $G \approx G^*$, or $G \prec G^*$ – will fail for certain (G, G^*) pairs. Whenever it does, there will be wagers that lack fair prices; that is, for some $G \in \mathbf{G}$ the agent's preference ranking will not contain *any* indifference of the form $\$x \approx G$, for x a real number. In such cases, one can still employ talk of G 's fair price to convey information about the agent's preferences, using things like, " G 's fair price lies between \$8,000 and \$9,000," to mean that she strictly prefers (disprefers) G to every sum less (greater) than \$8,000 (\$9,000). In speaking this way, however, we must be careful not to slip into what Isaac Levi calls "black box" thinking, which treats any inexact statement about a wager's fair price for a gambler as nothing more than an expression of our ignorance of its true price for her.³⁰ In cases where the strengths of her desires are genuinely indeterminate there will be wagers whose fair price we cannot know, not because our access to her mental states is somehow limited, but because there is nothing to be known. More generally, we must take care not to regard the incompleteness in a rational gambler's preference ranking as indicating gaps in our (or her) knowledge about the true interval scale that measures the strengths of her desires. A gambler can be rational even if her beliefs are too vague to be measured on any single interval scale. We must, therefore, reject the idea that the von Neumann/Morgenstern axioms are all necessary conditions for instrumental rationality since such a view would falsely entail that there can be no vagueness in preferences.

The problematic axiom is the completeness principle VN_6 . Many proponents of expected utility theory³¹ now interpret this axiom not as a law of rationality per se, but as a *requirement of coherent extendibility*. We will discuss the implications of this at some length in Chapter 3. The basic idea is that VN_6 should be replaced by

VN_6^+ *Coherent Extendibility*: While (\succ , \succeq) need not completely order \mathcal{G} , there must be at least one (complete) preference ranking (\succ^+ , \succeq^+) that both satisfies the axioms $\text{VN}_1 - \text{VN}_6$ and *extends* (\succ , \succeq) in the sense that $G \succ G^*$ implies $G \succ^+ G^*$ and $G \succeq G^*$ implies $G \succeq^+ G^*$.

In other words, it should always be possible in principle to “flesh out” a rational agent’s preferences (usually in more than one way) to obtain a complete ranking that does not violate any of the von Neumann axioms.

When VN_6 is replaced by VN_6^+ the expected utility theorist’s basic requirement of rationality becomes

Subjective Expected Utility Hypothesis (weak version). In order for a “professional” gambler to be rational in the instrumental sense there must be at least one utility function for money, u , whose associated expectation operator, U , *weakly ordinally represents* the gambler’s preference ranking in the sense that

$$\begin{aligned} G \succ G^* &\text{ only if } U(G) > U(G^*) \text{ and} \\ G \succeq G^* &\text{ only if } U(G) \geq U(G^*) \end{aligned}$$

hold for any $G, G^* \in \mathcal{G}$.

This, finally, is the principle that expresses the basic truth underlying Pascal’s insight: It sets up *representation of desires by some utility function or other* as the basic criterion of instrumental rationality. It does not, however, ask that there be a single utility that correctly measures the strength of an agent’s desires (up to the choice of a unit and zero point), nor does it force the agent to have a definite fair price for every wager she considers. Nonetheless, it preserves the essential core of the Pascalian position by making expected utility maximization central to rational betting, and by retaining the idea that there should always be at least one (but usually many) utility functions for which it is true $u(\$G) = U(G)$ in any case in which an agent’s desires are sufficiently definite to determine that $\$G$ is her fair price for G .

1.7 LIFE OUTSIDE THE CASINO

In my view, the weakened version of the Subjective Expected Utility Hypothesis expresses a correct norm of instrumental rationality for “professional” bettors who are evaluating risky wagers. This is, of course, a rather limited claim; it is only applicable to decisions made in the casino when money is at stake. The authors of the *Port Royal Logic* would not be pleased with us if we stopped here. They saw the new concepts of probability and expectation not merely as gambling aids, but as the foundation for a general methodology of right decision making. The final section of the *Logic*, in which the material on probability and expectation is found, has a distinctly apostolic ring to it. Its author, Antoine Arnauld (who either consulted Pascal directly or worked from notes

left after his death in 1662), clearly felt that Pascal was on to something much more important than a recipe for successful gambling. He saw himself as presenting an approach to decision making that could be applied everywhere to improve people's lives. This is nowhere explicitly stated, but the examples tell the tale. An appreciation of the laws of probability and expectation is supposed to liberate the ignorant from irrational fears of being struck by lightning or other similar events that, though awful in their consequences, have such low probabilities that it is not worthwhile to worry about them. It will also help foolish princesses overcome their fear of walking into rooms for fear that the ceiling will fall on them. It can even be used to make the biggest decision of all, whether to take actions that would lead to a belief in God, which Pascal famously viewed as a kind of a wager.

Thus, the core tenet of utility theory was already in place more than four-hundred years ago in the form of Pascal's Thesis that *rational decision making is a matter of using probabilities to calculate expected values for risky or uncertain prospects and of letting these expected values guide one's choices. Rational choice, in short, is governed by the laws of mathematical expectation.* There have been significant changes in decision theory since Pascal's time, but this basic insight remains at the core of all subsequent work.

The goal of future chapters is to see how this insight can be generalized and expanded to include more realistic decision making situations, such as those in which the "prizes" to be won are not merely monetary, and those in which the decision maker does not know the objective probabilities of all the events that are relevant to the outcomes of her choices. We need to start by getting clear about what decisions are.

¹ Arnauld and Nicole (1996, pp. 273-74).

² Arnauld and Nicole (1996, pp. 274-75).

³ The best discussion of the Pascal/Fermat correspondence is found in Todhunter (1865/1949, pp. 7-21).

⁴ While the idea that wagers should be valued by their expected payoffs had occurred to Pascal as early as 1654, it was Christian Huygens who first defended it in print in his immensely influential 1659 textbook *De Ratiociniis in Ludo Aleoe*.

⁵ For a discussion of the research in this area see Shafir and Tversky (1995).

⁶ Notice that the expression “\$g” is doing double duty here. It signifies both the sum of money *g dollars*, and the proposition *that the agent’s total fortune is g dollars*. This ambiguity should cause no confusion since the context will always determine which of the two is meant.

⁷ For readers interested in the details, I recommend Royden (1968).

⁸ The classic reference here is Savage (1954/1972).

⁹ See Ramsey (1931), de Finetti (1964), von Neumann and Morgenstern (1953), and Savage (1954/1972).

¹⁰ The mixture-as-compound view also has the unfortunate effect of leading people to think that mixtures have a temporal structure. One tends to look at H_1 and think, “First the nickel is tossed to decide whether the gambler gets G_1 or G_2 ; then the dime is tossed to determine her fortune.” For a professional bettor the sequence of the events that

instantiate a given wager is immaterial to its desirability except insofar as it determines the probabilities with which she will receive various fortunes.

¹¹ In speaking this way we must not be misled into thinking that the ability to buy happiness is what *makes* having money desirable for professional gamblers. The order is reversed. Since a professional gambler takes money to be a *basic* good, it buys her happiness because she desires it, *not* the other way around. This is what makes “professional” gamblers so unlike you and me.

¹² Substitution follows from the concept of a fair price via two applications of Independence. For the rest one may consult Savage (1954/1972, pp. 70-73) or Kreps (1988, pp. 43-51).

¹³ For example: Fishburn (1970), Savage ([1954] 1972), Kreps (1988).

¹⁴ The paradigmatic examples of interval scales are the Celsius and Fahrenheit measures of temperature. Each scale assigns “degrees” of temperature to physical objects in such a way that: the thermal ordering among of objects is invariant. While ratios of individual temperatures are not meaningful since they vary with the choice of scale, ratios of temperature differences are cardinally significant. So, while it makes sense to claim that an object at 100°F (= 37.38°C) is *hotter than* one at 50°F (= 10°C), it does not make sense to say that the first is twice (or 3.738 times) as hot as the second. One can, however, say that the *temperature difference* between a body at 50°F and one at 100°F is twice as great as the temperature difference between a body at 25°F (= -3.89°C) and one at 50°F

because the ratio $[100^{\circ}\text{F} - 50^{\circ}\text{F}]/[50^{\circ}\text{F} - 25^{\circ}\text{F}] = [37.38^{\circ}\text{C} - 10^{\circ}\text{C}]/[10^{\circ}\text{C} - (-3.89^{\circ}\text{C})] = 2$ is invariant among all legitimate ways of measuring temperature.

¹⁵ For an excellent discussion of the history see Todhunter (1865/1949, pp. 213-238).

¹⁶ A determined proponent of the Thesis might try to respond by suggesting that the (fair price = expected payoff) equation only applies to wagers based on *finite* event partitions or, at most to infinite wagers with expected payoffs that converge to some finite value.

This dodge will not work. The paradoxical character of the St. Petersburg game has nothing special to do with the fact that its expected payoff is infinite. The advice Pascal's Thesis offers is just as absurd in a "truncated" St. Petersburg game that has a finite upper limit of, say, one hundred thousand tosses (where it is stipulated that a bettor's fortune stays at $\$f$ if all one hundred thousand come up heads).

¹⁷ I am not using *happiness* here as the name of a pleasurable or satisfied *feeling*, but in the broader sense in which happiness encompasses all aspects of (subjective) well-being, the way the term is used in "1987 was the happiest year of my life."

¹⁸ This might not be obvious. Since the terms of the series $\sum_k 1/2^k u(2^k)$ can be chosen to be positive one can appeal to a version of the *ratio test*: If $a_k > 0$ for $k = 1, 2, 3, \dots$, then $\sum_k a_k$ converges if $a_{k+1}/a_k < 1$ for all k . Applying this with $a_k = u(2^k)/2^k$ yields the desired result.

¹⁹ There are few economic hypotheses that have more explanatory power than the declining marginal utility of money. It accounts for the willingness of poor people to work harder for a dollar than most rich people would ever dream of; the tendency of the

wealthy to spend a higher fraction of their income on “luxuries” and their willingness to invest more heavily in risky ventures (including education). It also explains the existence of institutions such as the insurance industry, trading in common stocks and futures, sharecropping, sweatshops, and long term labor contracts.

²⁰ This Nicholas is not to be confused with Daniel’s *brother* Nicholas, who posed the St. Petersburg problem to Cramer.

²¹ Proof: $[1/2\log(f + \delta) + 1/2\log(f - \delta)] = 1/2\log(f^2 - \delta^2) < 1/2\log(f^2) = \log(f)$.

²² I first heard this criticism of Pascal’s wager (albeit in a slightly different form) from Alan Hajek. His (19xx)” is the most sophisticated treatment of the wager I have yet seen. The basic point had been made earlier by Duff in his (1986).

²³ This can be done consistently. For details see Kreps (1988, pp. 63-68).

²⁴ I am taking this use of the term *constraint* from van Fraassen (1984).

²⁵ The terminology I am using is somewhat non-standard. Usually the phrase “preference ranking” is taken to denote a *complete ordering* of \mathbf{G} in which the “no preference” alternative never occurs, so that the *trichotomy property* – $G \succ G^*$ or $G \approx G^*$ or $G \prec G^*$ – holds for every pair $G, G^* \in \mathbf{G}$. However, since we are going to need a convenient term to denote both incomplete order relations, or *partial orderings*, as well as complete orderings, I am going use “ranking” to refer to both. This should cause no confusion so long as readers keep in mind that *rankings need not be complete*.

²⁶ The idea of using utilities to represent preferences had appeared earlier in the writings of Frank Ramsey, but his work was not well known, and not very influential, until after the publication of *Theory of Games*.

²⁷ I have altered the von Neumann/Morgenstern axioms merely to highlight the analogies with Premises 1-7.

²⁸ Kaplan (1994, pp. 23-31)

²⁹ A bad reply here would be to suggest that your utility for \$100,000 just *is* that number you would *choose* if you had to pick a value for ρ at which you would get either the money or the wager with equal probability. Aside from taking us back to the behaviorist view of desire rejected above, this assumes, falsely, that the choice you make will always reveal your degree of desire. It probably would if you had one, but it is perfectly consistent with having no definite degree of desire that you choose a precise value for ρ when forced to do so. This is especially evident when the choice is not underwritten by any *stable* behavioral disposition.

³⁰ Levi (1985).

³¹ See, for example, Kaplan (1989) and (1996), and Jeffrey (1983).