ideas they comprehend as quickly and automatically as they believe in the objects they see." At the same time our belief system has mechanisms that allow for subsequent unacceptance of ideas. Within an evolutionary framework, Gilbert's claims and supporting findings make sense. However, outside of this framework they may seem counterintuitive.

An Important Moral

It is clear that researchers must be very careful about relying on their intuitions in formulating theories of thought and behavior. Experimental methods are absolutely essential for determining the validity of such intuitions. Perhaps less obvious is that familiarity and direct experience with psychological research may lead to better intuitions about thinking and behavior. Cognitive and social psychologists, like other scientists, attempt to develop theories that explain a wide range of phenomena and that predict new phenomena. Consider a psychologist who has developed a theory that explains a number of surprising and counterintuitive findings. The psychologist will use the theory to make predictions about new phenomena that will be intuitive to the psychologist (because they follow from the theory) but that are likely to be surprising and counterintuitive to researchers and lay people who do not know about the theory. To the extent that the theory is a good one (by the usual scientific criteria), the psychologist's intuitions are more likely to turn out right than those of people who are not familiar with the theory and the phenomena it explains.

Chapter 4

Philosophical Intuitions and Cognitive Mechanisms

Eldar Shafir

Intuition occupies a central role in philosophical theorizing. Some of the most poignant and memorable passages in philosophical writings have relied on examples whose appeal to intuition can make compelling a theory that until then seemed obscure. The appeal to intuition can be observed in domains ranging from metaphysics and epistemology, to ethics and the philosophy of mind. In what follows, I shall be unabashedly descriptive in my treatment of intuitions. I shall focus on systematic and well-documented aspects of the psychology that underlies people's intuitions; I shall ignore questions such as whether there are moral facts, or facts about rationality, and whether we may have intuitive, perceptual, or other privileged access to such facts. This chapter will consider the systematic ways in which intuitions shift as a result of supposedly inconsequential manipulations, and the implications this might have for the stability and significance of philosophical theorizing.

A descriptive account of the psychology that underlies people's attitudes and intuitions should be given serious consideration, even by scholars mostly concerned with normative or prescriptive theory. The compelling nature of normative theory notwithstanding, most scholars of human behavior tend to endorse theories that they consider psychologically feasible. Even those who suppose an exceedingly high degree of rationality or morality on the part of individuals have typically regarded their assumptions to be plausible, if somewhat idealized. Unwilling to deny the relevance of human nature, these theorists adopt a naive account of mental life that, if approximately correct, could yield behaviors largely consistent with those dictated by normative theory. Requirements of deductive closure or unbounded memory, for example, are obviously unrealistic about us and thus not part of the assumptions that most people make. Likewise, moral principles are taken seriously to the extent that the creatures to which they are applied are assumed to be able to follow them. Many errors of reasoning, inconsistencies in choice, failures of self-control, and moral transgressions, to name a few, are considered interesting, if not embarrassing, precisely because there is the feeling that one could, and should, have done better.

The descriptive approach is based on empirical observation and experimental studies of behavior. The evidence indicates that people's sentiments and preferences exhibit patterns that are often at odds with intuitive assumptions, and em-

pirical generalizations emerge that help explain the nonintuitive patterns. In what follows, I review selected findings and discuss some psychological principles that underlie preference and evaluation. In particular, I focus on a systematic discrepancy that is observed between evaluations that are conducted in isolation, when one alternative is considered at a time, and choices that are observed in comparative settings, when two or more alternatives are considered simultaneously. Typically, isolated evaluations are obtained in a "between-subject" design, where some people evaluate one scenario and others evaluate another, or when the same person evaluates different scenarios sequentially, at different points in time, so as to render direct comparison difficult. Simultaneous evaluations are observed in a "within-subject" design, when a person is presented with two or more scenarios concurrently. The systematic discrepancy observed between the two modes of evaluation has profound implications for the role of philosophical intuition. Whereas most life experiences take place in what can be thought of as a between-subject design (you encounter one scenario; someone else may encounter another), philosophical intuitions typically are the introspective result of a within-subject evaluation (a philosopher contemplates a scenario and its alternatives). The implications of this tension are explored in what follows (see also Kahneman 1996). In the next two sections, alternative elicitation methods are shown to give rise to differential weighting of dimensions and, consequently, to inconsistent decisions. Related phenomena are then reviewed in the realm of counterfactual evaluation and in contexts that explore people's feelings of sympathy, urgency, and indignation. Section IV contrasts the phenomenology of uncertainty with compelling intuitions about reasoning in uncertain situations. Concluding remarks occupy the last section.

Compatibility and Preference Reversals

Elicitation of Preference

Preferences can be elicited through different methods. People can indicate which option they prefer; alternatively, they can be asked to price each option by stating the amount of money that is as valuable to them as the option. A standard assumption, known as *procedure invariance,* requires that logically equivalent elicitation procedures give rise to the same preference order. Thus, if one option is chosen over another, it is also expected to be priced higher. Procedure invariance is essential for the interpretation of both psychological and physical measurement. The ordering of physical objects with respect to mass, for example, can be established either by placing each object separately on a scale, or by placing both objects on two sides of a pan balance. Procedure invariance requires that the two methods yield the same ordering, within the limit of measurement error. Analogously, the rational theory of choice assumes that an individual has a well-defined preference order that can be elicited either by choice or by pricing, giving rise to the same ordering of preferences.

Compatibility Effects

Despite its appeal as an abstract principle, people systematically violate procedure invariance. For example, people often choose one bet over another, but price the second bet above the first. In one study, subjects were presented with two prospects of similar expected value. One prospect, the H bet, offered a high probability to win a relatively small payoff (e.g., 8 chances in 9 to win $4) whereas the other prospect, the L bet, offered a low probability to win a larger payoff (e.g., a 1 in 9 chance to win $40). When asked to choose between these prospects, most subjects chose the H bet over the L bet. Subjects were also asked, on another occasion, to price each prospect by indicating the smallest amount of money for which they would be willing to sell this prospect. Here, most subjects assigned a higher price to the L bet than to the H bet. One recent study that used this particular pair of bets observed that 71 percent of the subjects chose the H bet, while 67 percent priced L above H (Tversky, Slovic, and Kahneman 1990). This phenomenon, called "preference reversal," has been replicated in experiments using a variety of prospects and incentive schemes; it has been observed in a study involving professional gamblers in a Las Vegas casino (Lichtenstein and Slovic 1973; Slovic and Lichtenstein 1983), and in a study conducted in the Peoples' Republic of China for real payoffs equal to several months' worth of the subjects' salary (Kachelmeier and Shehata 1992).

What is the cause of preference reversal? Why do people assign a higher monetary value to the low probability bet, but choose the high probability bet more often? It appears that the major cause of preference reversal is a differential weighting of probabilities and payoffs in choice and pricing, induced by the type of response. In line with the general notion of compatibility, which has long been recognized by students of perception and motor control, experimental evidence indicates that an attribute of an option is given more weight when it is compatible with the response format than when it is not (Tversky, Sattath, and Slovic 1988; for review, see Shafir 1995). Because the price that the subject assigns to a bet is expressed in dollars, the payoffs of the bet, which are also expressed in dollars, are weighted more heavily in pricing than in choice. As a consequence, the L bet (which has the higher payoff) is evaluated more favorably in pricing than in choice, which can give rise to preference reversals. (The foregoing account is further supported by the observation that the incidence of preference reversals is greatly reduced for bets involving nonmonetary outcomes, such as a free dinner at a local restaurant, where outcomes and prices are no longer expressed in the same units and are therefore less compatible; see Slovic, Griffin, and Tversky 1990.)

The compatibility hypothesis does not depend on the presence of risk. It predicts a similar discrepancy between choice and pricing in the context of riskless options that have a monetary component. Consider a long-term prospect L, which pays $2,500 five years from now, and a short-term prospect S, which pays $1,600 in one and a half years. Subjects were invited to choose between L and S and to price both prospects by stating the smallest immediate cash payment for which they would be willing to exchange each prospect (Tversky, Slovic, and Kahneman 1990). Because the payoffs and the prices again are ex-

pressed in the same units, compatibility suggests that the long-term prospect (offering the higher payoff) will be overvalued in pricing relative to choice. In accord with this hypothesis, subjects chose the short-term prospect 74 percent of the time but priced the long-term prospect above the short-term prospect 75 percent of the time. These observations indicate that different methods of elicitation, such as choice and pricing, can induce different weightings of attributes, which, in turn, can give rise to different preferences.

Note that the pricing of an option involves independent evaluation and could be used to assign worth in isolation. Choice, on the other hand, is an inherently comparative process, that requires concurrent presentation. When options are encountered one at a time, they can be priced or assigned other measures of attractiveness, but direct comparison is not feasible. The weights that enter into our evaluations when conducted in isolation are thus expected to differ from those that characterize concurrent evaluation, when both options are before the attention. When a philosopher introspects about how people will, or even ought to, evaluate different options when these are presumably encountered in isolation, the philosopher will be confined to a concurrent evaluation, with the various alternatives before his or her attention. To the extent that the two forms of evaluation—concurrent and in isolation—lead to differential weightings, there will be a systematic tendency for people to experience events in isolation that will remain beyond the scope of within-subject intuition. In everyday life, we tend to experience and evaluate scenarios one at a time, whereas intuitions about relative worth often arise from introspectively comparative evaluations. As a result, the weighting of dimensions that goes into the making of common intuition can differ systematically from the weights that are assigned in actual experience.

The Prominence Hypothesis and Reversals in Perceived Importance

People often feel that one attribute (e.g., safety) is more important than another (e.g., cost). Although the interpretation of such claims is not entirely clear, there is evidence that the attribute that is judged more important looms larger in choice than in independent evaluation, such as pricing (Slovic 1975; Tversky, Sattath, and Slovic 1988). This is known as the prominence hypothesis and can lead to systematic violations of invariance.

Consider, for example, the following study concerning people's responses to environmental problems (Kahneman and Ritov 1994). Several pairs of issues where selected, where one issue addresses human health or safety and the other concerns protection of the environment. Each issue included a brief statement of a problem, along with a suggested form of intervention, as illustrated below.

> *Problem*: Skin cancer from sun exposure is common among farm workers.
> *Intervention*: Support free medical checkups for threatened groups.

> *Problem*: Several Australian mammal species are nearly wiped out by hunters.
> *Intervention*: Contribute to a fund to provide safe breeding areas for these species.

One group of respondents was asked to choose which of the two interventions they would rather support; a second group was presented with one issue at a time and asked to determine the largest amount they would be willing to pay for the respective intervention. Because the treatment of cancer in humans is generally viewed as more important than the protection of Australian mammals, the prominence hypothesis predicts that the former will receive greater support in direct choice than in independent evaluation. This prediction was confirmed. When asked to evaluate each intervention separately, respondents, who might have been moved by these animals' plight, were willing to pay more, on average, for safe breeding of Australian mammals than for free checkups for skin cancer. However, when faced with a direct choice between these options, most subjects favored free checkups for humans over safe breeding for the mammals. As expected, the issue that is considered more important acquired a greater prominence in the choice condition, which allows for a direct comparison between issues, than in separate presentation, where each issue is evaluated in accord with its own generated emotions. Irwin, Slovic, Lichtenstein, and McClelland (1993) report related findings in settings where improvements in air quality were compared with improvements in consumer commodities. In general, people may evaluate one alternative more positively than another when these are evaluated independently, but then reverse their evaluation in direct comparisons that accentuate the prominent attribute.

A similar pattern may occur in cases where an attribute is particularly difficult to gauge in isolation. Hsee (1996), for example, presented subjects with two alternative second-hand music dictionaries; one with 20,000 entries but a slightly torn cover, the other with 10,000 entries and a cover like new. Subjects had little notion of how many entries to expect in a music dictionary. Consequently, under separate evaluation, they expressed a willingness to pay more for the dictionary with the new cover than for the one with a slightly torn cover. When the two dictionaries were evaluated concurrently, however, most subjects obviously preferred the dictionary with twice as many entries, despite its inferior cover.

Intuitions about importance, worth, gravity, as well as ethical propriety are often obtained in comparative settings; we ask ourselves which issue, A or B, is more grave, or more worthy of our attention; which act, A or B, constitutes a greater ethical violation. In life, we often encounter the relevant scenarios one at a time; we might encounter scenario A today, and somebody else, or we, at another time, might encounter scenario B. To the extent that our encounters with these scenarios trigger sentiments and reactions that partly depend on their being experienced in isolation, some critical (and perhaps normatively appropriate) aspects of our response are likely to be missed by intuitions that arise from concurrent, within-subject introspection.

Affect and Principles

In a study ostensibly intended to establish the amounts of compensation payment that the public considers reasonable, Miller and McFarland (1986) presented respondents with brief descriptions of victims who had applied for compensation and asked them to decide upon a monetary payment. Two such descriptions concerned a male victim who was described as having lost the use of his right arm as a result of a gunshot wound suffered during a robbery at a convenience store. Some respondents were told that the robbery happened at the victim's regular store. Others were told that the victim was shot at a store he rarely frequented, that he went to because his usual store was temporarily closed. It was hypothesized that subjects would assign higher compensation to a person whose victimization was preceded by an abnormal event. This is because abnormal events strongly evoke a counterfactual undoing, which tends to raise the perceived poignancy of outcomes and the sympathy for their victims. (For more on the psychology of counterfactual thinking, see Kahneman and Miller 1986; Roese and Olson 1995.) Indeed, the victim who was shot at a store he rarely visited was assigned significantly greater compensation than the victim who was shot at his regular store. The difference in poignancy created by the normal-versus-abnormal manipulation translated into a $100,000 difference in compensation judged appropriate for the two cases.

The affective impact of events is often influenced by the ease with which an alternative event can be imagined. The death of a soldier on the last day of the war seems more poignant than the death of his comrade six months earlier. The fate of a plane crash victim who switched to the fatal flight only minutes before take-off is seen as more tragic than that of a fellow passenger who had been booked on that flight for months. Whereas the affective impact of such distinctions is predictable and often strong, do people actually consider these distinctions relevant? Consider the earlier study about compensation to victims. Recall that the two versions of the robbery scenario—at the regular versus the unusual stores—were presented to separate group of subjects. Their affective responses—stronger for the unusual than for the regular scenario—were thus obtained in isolation. On the other hand, when respondents were presented with both versions concurrently, the great majority (90 percent) thought that the victims in the two cases should not be awarded different compensations (Kahneman 1996). Evidently, despite the large difference in awards observed above, most subjects consider the difference between the two scenarios irrelevant to compensation. In a within-subject design that allows direct comparison, rules about what is relevant are easy to apply: we can decide, for example, that the victim's past frequency of visits to the store is immaterial. Between subjects, on the other hand, the application of rules remains elusive: there is no way to assure that the affective reactions that guide our response in isolation conform to the rules that would be endorsed upon concurrent evaluation. Using data from the 1992 Summer Olympics, Medvec, Madey, and Gilovich (1995) showed that athletes who had won silver medals tended to be less satisfied than those who won bronze. Apparently, the silver medalists compare themselves to those who had won the gold,

medals. Of course, if they had to choose all these athletes would presumably prefer the silver over the bronze. Thus, the feelings of relief or disappointment that loom large in the separate experiences are clearly overwhelmed by preference for a better placement upon concurrent evaluation.

The intensity of satisfaction, empathy, or indignation that we feel can be affected by nuanced factors. Principles of decision intended to transcend some of these factors can be compelling in direct comparisons, but difficult to apply in isolated evaluations. This tension presents interesting philosophical questions. In one study (Tversky and Griffin 1991), respondents were presented with two hypothetical job possibilities, one offering a higher yearly salary in a company where others with similar training earn more (You: $35,000; Others: $38,000), and the other offering a lower salary in a company where others with similar training earn less (You: $33,000; Others: $30,000). Most of us tend to abide by a simple principle according to which we ought to prefer outcomes that improve our lot more over outcomes that improve it less. In fact, a majority of respondents chose the job with the higher absolute salary, despite the lower relative position. This simple principle, however, does not apply with equal force when we contemplate each of the job offers separately: in this condition, without the other offer serving as a comparison, earning a salary lower than comparable others can highlight sentiments that reduce our feelings of satisfaction. Indeed, contrary to the preference observed above, the majority of respondents who evaluated each of the job offers separately anticipated higher satisfaction in the job with the higher relative position and lower salary. A variant of this study was replicated with second-year MBA students, who were presented with two alternative job offers. In one, they would be paid $75,000, the same as other starting MBAs; in the other, they would be paid $85,000 while some other graduating MBAs would receive $95,000. As predicted, the students proved more willing to accept the former job offer when these were evaluated in isolation, but chose the latter offer when the two were evaluated concurrently (Bazerman et al. 1994; see also Bazerman, White, and Loewenstein 1995, for related discussion).

It is interesting to note in this context that consequentialist or utilitarian considerations appear to loom larger in concurrent than in isolated evaluations. In line with related observations regarding the malleability of utility estimation in decision making, it seems that the utilitarian worth of outcomes, which is often hard to gauge out of context, plays a greater role in direct comparisons than in isolated settings. Hsee (1997), for example, presented subjects with pictures of two servings of Hägen-Dazs ice cream. One serving contained more ice cream that failed to fill a larger cup; the other contained less ice cream that overfilled a smaller container. When the two were evaluated jointly, subjects were willing to pay more for what was clearly a larger serving. In separate evaluation, however, when the precise amount of ice cream was hard to gauge, subjects tended to pay more for the overfilled cup than for the one that seems partly empty.

Simple principles of merit, entitlement, worth, or maximization, which can play a decisive role in comparative settings, often prove difficult to apply in isolated situations. The compensation a victim is entitled to, the attractiveness of a job offer, or the value of a serving of ice cream can be hard to gauge when these occur in isolation. In fact, other considerations, such as the emotional impact of

the victim's plight, the sense of fairness produced by a co-worker's salary, or the amount of ice cream relative to the size of the container, can strongly influence our evaluations when these occur in isolation. To the extent that our experiences with such matters generate sentiments and reactions that partly depend on their being evaluated in isolation, these important aspects of our affective responses are likely to be missed by intuitions that arise from well-defined principles that are sometimes only possible to apply in concurrent, within-subject evaluations.

Uncertainty and the Sure-Thing Principle

Many decisions are made in the presence of some uncertainty about their consequences. A critical feature of thinking under uncertainty is the need to consider possible states of the world and their potential consequences for our beliefs and actions. A fundamental principle which underlies most analyses of rational choice was described by Savage (1954: 21), who captured the intuition in the following passage:

> A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant to the attractiveness of the purchase. So, to clarify the matter for himself, he asks whether he would buy if he knew that the Republican candidate were going to win, and decides that he would do so. Similarly, he considers whether he would buy if he knew that the Democratic candidate were going to win, and again finds that he would do so. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains. . . . It is all too seldom that a decision can be arrived at on the basis of the principle used by this businessman, but, except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance.

Savage went on to define this principle formally: If x is preferred to y knowing that event A obtained, and if x is preferred to y knowing that event A did not obtain, then x should be preferred to y even when it is not known whether A obtained. As Savage points out, this principle, which he called the *sure-thing principle* (henceforth, STP), has a great deal of both normative and descriptive appeal. It is one of the simplest and least controversial principles of rational behavior and is implied by "consequentialist" accounts of decision making, in that it captures a fundamental intuition about what it means for a decision to be determined by the anticipated consequences.[1] It is a cornerstone of Expected Utility Theory, and it holds in other models of choice that impose less stringent criteria of rationality. It is intuitively very compelling. Nonetheless, people's decisions do not always abide by STP.

The Disjunction Effect

Consider the following problem that occurs in one of two versions, as indicated

> Imagine that you have just played a game of chance that gave you a 50 percent chance to win $200 and a 50 percent chance to lose $100. The coin was tossed and you have [won $200 / lost $100]. You are now offered a second, identical gamble: 50 percent chance to win $200 and 50 percent chance to lose $100. Would you:

	Won	Lost
a) Accept the second gamble	69%	59%
b) Reject the second gamble	31%	41%

Tversky and Shafir (1992) presented subjects (ninety-eight Stanford undergraduates) with the Won version of the problem above, followed a week later by the Lost version, and ten days after that by the following version that is a disjunction of the previous two:

> Imagine that you have just played a game of chance that gave you a 50 percent chance to win $200 and a 50 percent chance to lose $100. Imagine that the coin has already been tossed, but that you will not know whether you have won $200 or lost $100 until you make your decision concerning a second, identical gamble: 50 percent chance to win $200 and 50 percent chance to lose $100. Would you:

a) Accept the second gamble	36%
b) Reject the second gamble	64%

These problems were embedded among several others and temporally separated so the relation among the three versions was not transparent. To the right of each option is the percentage of subjects who chose it. The data show that the majority of subjects accepted the second gamble after having won the first gamble, and the majority accepted the second gamble after having lost the first gamble. However, contrary to STP, the majority of subjects rejected the second gamble when the outcome of the first was not known. Among those subjects who accepted the second gamble, both after a gain and after a loss on the first, 65 percent rejected the second gamble in the disjunctive condition, when the outcome of the first gamble was uncertain. In fact, this particular pattern—accept when won, accept when lost, but reject when do not know—was the single most frequent pattern exhibited by these subjects (see Tversky and Shafir 1992; Shafir and Tversky 1992, for more data and discussion). We call this pattern a *disjunction effect.* A disjunction effect occurs when a person prefers x over y when she knows that event A obtains, and she also prefers x over y when she knows that event A does not obtain, but she prefers y over x when it is unknown whether or not A obtains. The disjunction effect amounts to a violation of STP, and hence of consequentialism.

When confronted with the disjunctive scenario above, our subjects appear not to evaluate the attractiveness of the

tions, one assuming a gain and one assuming a loss, as implied by STP. Instead, the presence of uncertainty induces its own phenomenology, in which the unresolved outcome looms large and unknown. The first gamble has a positive expected value, but it also involves the risk of a nontrivial loss. In the Won condition, the decisionmaker is already up $200, so regardless of the outcome of the second gamble, he is assured to remain ahead overall, which makes the gamble quite attractive. In the Lost condition, the decisionmaker is down $100: since most people hate a sure loss, the second gamble offers an attractive chance to "get out of the red." In the disjunctive condition, however, neither motive is entirely compelling. The decisionmaker experiences neither the reassurance that comes with knowing that he can no longer lose, nor the compulsion to recover recent losses; instead, a prevalent attitude is one of caution, a reluctance to rush into further action when previous ones have not yet been resolved. (For related analyses in terms of reasons in choice, see Shafir, Simonson, and Tversky 1993.)

We have replicated the above effect in a between-subject design. Three different groups of subjects were presented with the Won version, the Loss version, and the disjunctive version. As with the previous study, a majority accepted the gamble in the Won and in the Loss conditions (69 percent and 57 percent, respectively), but only 38 percent accepted it in the disjunctive condition. The fact that the distribution of choices was nearly identical in the two studies suggests that the respondents in the original study evaluated each version independently, with no detectable effects of one version on another. In fact, although technically a "within-subject" design, the original study obtained clearly independent evaluations, thus rendering it comparable to a between-subject manipulation.

A Theoretical Analysis

The above disjunction effect may be interpreted in terms of the value function from Kahneman and Tversky's (1979) Prospect Theory. The function, shown in figure 4.1, represents the subjective value of modest gains and losses and has been generally supported by numerous empirical studies. In accord with the principle of diminishing sensitivity, the function incorporates a concave segment to the right of the origin, namely, in the domain of gains, and a convex segment to the left, in the domain of losses. Furthermore, the function is steeper for losses than for gains, in accord with the principle of loss aversion.[2] The function in figure 4.1 represents a typical decisionmaker who is indifferent between a 50 percent chance of winning $100 and a sure gain of roughly $35, and, similarly, is indifferent between a 50 percent chance of losing $100 and a sure loss of roughly $40. Such a pattern of preferences can be captured by a power function with an exponent of .65 for gains and .75 for losses.

Consider, then, a person P whose values for gains and losses are captured by the function of figure 4.1. Suppose that P is presented with the gamble problem above and is told that he has won the first toss. He now needs to decide whether to accept or reject the second. P needs to decide, in other words, whether to maintain a sure gain of $200 or, instead, opt for an equal chance at either a $100 or a $400 gain. Given P's value function, his choice is between two options whose

Accept the second gamble: $.50 * 400^{(.65)} + .50 * 100^{(.65)}$

Reject the second gamble: $1.0 * 200^{(.65)}$

Because the value of the first alternative is greater than that of the second, P is predicted to accept the second gamble. Similarly, when P is told that he has lost the first gamble and needs to decide whether to accept or reject the second, P faces the following options:

Accept the second gamble: $.50 * -[200^{(.75)}] + .50 * 100^{(.65)}$

Reject the second gamble: $1.0 * -[100^{(.75)}]$

Again, because the first quantity is larger than the second, P accepts the second gamble.

Thus, once the outcome of the first gamble is known, the function in figure

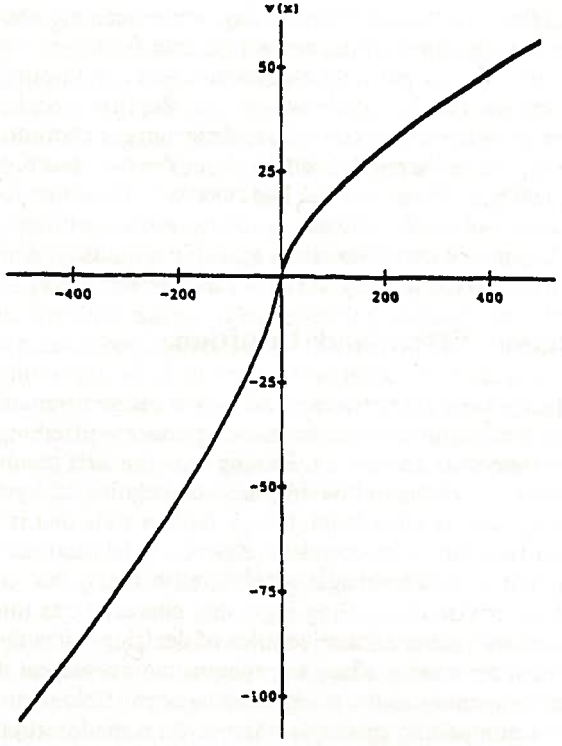

Figure 4.1
The value function $v(x) = x^{.65}$ for $x \geq 0$ and $v(x) = -(-x)^{.75}$ for $x \leq 0$.

4.1 predicts that person P will accept the second gamble whether he has won or lost the first. But as long as the outcome of the first gamble is not known, P might proceed as if for the moment no change has transpired. Not knowing whether he has won or lost, P assumes that he is still where he began, at the status quo, at the origin of his value function. When faced with the decision to accept or reject the second gamble, P evaluates it from his original position, without aggregating the outcome of the first gamble, which remains unknown. Thus, P is deciding to accept or to reject a gamble that offers an equal chance to win $200 or lose $100:

Accept the second gamble: $.50 * -[100^{(.75)}] + .50 * 200^{(.65)}$

Reject the second gamble: 0

Because the expected value of accepting is just below 0, P decides to reject the second gamble in this condition.

In situations of uncertainty, different outcomes often do trigger different actions. It can be reasonable in such cases to suspend judgment until there is further resolution. When confronted with the disjunctive scenario above, people do not evaluate the attractiveness of the second gamble from two alternative positions, one assuming a gain and one assuming a loss, as implied by STP. Instead, not knowing whether they have won or lost the first, people segregate the two gambles and evaluate the second from their current position, as if for the moment no change has occurred. Uncertain about the previous outcome, people evaluate the situation as if no outcome had obtained. This interpretation is further supported by the observation that a similar percentage of people accept the gamble in the disjunctive condition as in a simple condition in which no prior gamble had been played (36 percent and 33 percent, respectively).

The Disjunction Effect and Intuition

The above analysis offers a positive as well as a negative account. The positive account suggests that disjunctive situations bring about a different psychological state from when outcomes are certain. Having won the first gamble assures the person of a no-loss situation, and having lost compels her to try to recover the losses. Uncertainty, on the other hand, brings about a state that is not a disjunction of the former two, but an independent tendency to be cautious and avoid further losses. Implicit in this is a negative account, namely, that subjects do not see through the otherwise compelling logic that characterizes this situation. In fact, as is the case with other normative rules of decision, once the applicability of STP is detected, for example in a transparent within-subject design, people typically find it compelling to the point of being irresistible. But as long as the applicability of a compelling principle has not been made salient, mental life abides by rules of its own, often in direct contradiction to the patterns that are endorsed by contemplative, within-subject intuition.

Recall that in the original study we presented subjects with the Won, Lost, and disjunctive versions each a week apart, so that the logical relation among the versions was not detected. In another study, subjects were presented with all three versions concurrently, on the same page, thus rendering the applicability of the sure-thing principle transparent. The percentages of subjects who accepted the second gamble in the Won condition (71 percent) and in the Loss condition (56 percent), when these were presented concurrently, were almost identical to those observed originally, when the versions were presented a week apart. On the other hand, the tendency to accept the second gamble in the disjunctive condition rose from 36 percent in the original, separated presentation, to 84 percent in the concurrent presentation. In fact, the proportion of subjects in the concurrent presentation who exhibited the pattern "accept when won, accept when lost, but reject when do not know" declined by more than 80 percent relative to the separated presentation. Once people realize that they would accept the second gamble regardless of the outcome of the first, they are compelled to accept it in the disjunctive condition.

Violations of STP are likely to be observed only when people have not considered the implications of the possible outcomes, and are likely to disappear in transparent, within-subject presentation. Indeed, numerous studies have shown that, contrary to Savage's businessman, subjects often refrain from partitioning a scenario or a category into their constituent events or subcategories. In the face of uncertainty, various intellectual, emotional, and motivational factors can influence perception, often quite independently of how the situation is perceived once the uncertainty is resolved. This can lead to violations of STP when the uncertain condition is considered in isolation, as typically occurs in a between-subject design. On the other hand, a within-subject design, in which people consider their preference and observe that it remains unchanged throughout, renders salient the compelling intuition underlying STP. But then how could Savage, having just contemplated the relevant outcomes in his example, intuit the potential STP violation? In fact, with the alternative versions of the problem immediately before his attention, Savage is experiencing precisely the concurrent presentation condition described above, and in that condition the logic of STP proves inescapable. Philosophical intuitions such as those articulated by Savage involve the philosopher serving as subject in what amounts to a within-subject introspection. People's experiences, on the other hand, typically occur in between-subject conditions. Those aspects of behavior that are confined to a between-subject analysis are likely to go undetected by within-subject intuitions.

Concluding Remarks

A number of psychological factors were considered that occasionally contribute to inconsistent sentiments, judgments, and preferences in isolated versus concurrent evaluations. First, different methods of elicitation, such as choice versus pricing, were seen to induce divergent weightings of attributes and thus give rise to inconsistent preferences. Next, dimensions that were considered more important, or harder to evaluate, were seen to acquire greater prominence in concurrent presentations, which allow for direct contrast, than in isolated presentation. Similarly, rules of decision that favor some factors over others were seen to

decisive role in direct comparisons, but proved difficult to apply in isolated evaluations. Finally, a phenomenology of uncertainty that was observed in isolated presentation was hard to capture in concurrent, within-subject, introspection. This collection of instances, it was suggested, mirrors a discrepancy between the nature of people's everyday experiences and the conditions that yield philosophical intuitions. In life, people typically experience and evaluate things one at a time, as in a between-subject design, whereas many of the relevant intuitions result from concurrent, within-subject introspection.

Intuition need not always arise from a purely concurrent mode of evaluation. In fact, a person may attempt to evaluate one alternative "in isolation" and then proceed to evaluate the second. However, this attempt at a sequential evaluation of isolated events is likely to prove difficult and of limited success, particularly when—as in Savage's STP—the desired intuition depends on the interaction, or comparison, of the disparate evaluations. Furthermore, even if one were successful at intuiting reactions to events in isolation, that would not resolve the conflict with intuitions that emerge under a concurrent evaluation.

Within-subject introspection, it turns out, provides a better account of people's intuitions than of their actual behavior. Many principles of ethics and rationality are compelling because they originate from strong intuitions that most of us share. When confronted with judgments or preferences that violate normative principles, people often wish to modify their behavior to conform with the principles. Evidently, people's behavior is often at variance with their own normative intuitions. In this sense, both normative and descriptive accounts capture important aspects of human competence: the first addresses reflective deliberation, whereas the second focuses on actual behavior. The two analyses, of course, are interrelated but they do not coincide. Often, people prefer to adhere to normative principles, but these sometimes conflict in nontrivial ways with tendencies that arise in specific situations. Thus, people generally agree that one should contribute to worthy causes and ought to refrain from lying, despite the fact they do not always do so. Similarly, people tend to accept the normative force of invariance, despite the fact that it is often violated in their actual choices. The distinction between normative and descriptive accounts is easier to intuit when it stems from notions such as self-interest or lack of self-control; it proves less intuitive when the violation of normative principles stems from the nature of cognitive operations.

Because intuitions can be very compelling, counterintuitive findings often need to be demonstrated in between-subject designs. Only in such contexts can we discover certain facts about our mental life that cannot be accessed by intuition. This has obvious implications for the study of philosophical problems (see also Goldman 1993a, for further discussion). Consider, for example, the intuitive distinction most of us feel between acts of omission and acts of commission. Or between intentional versus nonintentional acts. Or between different forms of allocation, distribution, and justice. In most of these cases, our intuitions arise from direct comparison and concurrent evaluation. It seems important to know to what extent these sentiments are maintained in a between-subject context, when evaluated in isolation. In light of the findings above, we should

policy implications. Imagine that some distinctions our intuition tells us are important disappear in between-subject evaluations, and that distinctions we did not previously entertain suddenly prove important. What should we do then? Should we strive for arrangements that improve things according to intuitions that emerge from concurrent evaluation, or should we instead, contrary to our intuitions, strive to create a world that ameliorates experiences in between-subject conditions? You can entertain both these possibilities or, perhaps, you should consider one and I the other.

Notes

This work was supported by U.S. Public Health Service Grant No. 1-R29-MH46885 from the National Institute of Mental Health, and has benefited from discussions with Daniel Kahneman.

1. The notion of consequentialism appears in the philosophical and decision theoretic literature in a number of different senses. See, e.g., Hammond 1988; Levi 1991; and Bacharach and Hurley 1991, for technical discussion. See McClennen 1983, for a critique. See also Shafir and Tversky 1992, for a discussion of nonconsequential decision making.

2. For more on Prospect Theory, see Kahneman and Tversky 1979; Tversky and Kahneman 1986. For recent extensions of Prospect Theory, see Tversky and Kahneman 1992. For more on loss aversion, see Tversky and Kahneman 1991. Prospect theory also incorporates a weighting function that replaces stated probabilities by some nonadditive measure. In the present treatment the weights coincide with stated probabilities. This is for the sake of simplification: it is not essential to the analysis.