Experiences of Collaborative Research

Daniel Kahneman
Princeton University

The author's personal history of the research that led to his recognition in economics is described, focusing on the process of collaboration and on the experience of controversy. The author's collaboration with Amos Tversky dealt with 3 major topics: judgment under uncertainty, decision making, and framing effects. A subsequent collaboration, with the economist Richard Thaler, played a role in the development of behavioral economics. Procedures to make controversies more productive and constructive are suggested.

The Collaboration With Amos Tversky

It was the spring of 1969, and I was teaching a graduate seminar at the Hebrew University of Jerusalem on the applications of psychology to real-world problems. In what turned out to be a life-changing event, I asked my younger colleague Amos Tversky to tell the class about what was going on in his field of judgment and decision making. Amos told us about the work of his former mentor, Ward Edwards, whose lab was using a research paradigm in which participants were shown two book bags filled with poker chips. The bags were said to differ in their composition (e.g., 70 red and 30 white chips or 30 red and 70 white chips). One bag was randomly chosen. The participant was given an opportunity to sample successively from it and was required to indicate after each trial the probability that it came from the predominantly red bag. Edwards had concluded from the results that people are *conservative Bayesians:* They almost always adjust their confidence interval in the proper direction, but rarely far enough.

A lively discussion developed around Amos's talk. The idea that people were conservative Bayesians did not seem to fit with the everyday observation of people commonly jumping to conclusions. It also appeared unlikely that the results obtained in the sequential sampling paradigm would extend to the situation, arguably more typical, in which evidence is delivered all at once. Finally, the label *conservative Bayesian* suggested a process that first gets the correct answer, then adulterates it with a bias—not a plausible psychological mechanism. I learned recently that one of Amos's friends met him that day and heard about our conversation, which Amos described as having severely shaken his faith in the neo-Bayesian approach. Amos and I decided to meet for lunch to discuss our hunches about the manner in which probabilities are "really" judged. There we exchanged personal accounts of our own recurrent errors of judgment in this domain and decided to study the statistical intuitions of experts.

I spent the summer of 1969 doing research at the Applied Psychological Research Unit in Cambridge, England. Amos stopped there for a few days on his way to the United States. I had drafted a questionnaire on intuitions about sampling variability and statistical power, which was based largely on my personal experiences of incorrect research planning and unsuccessful replications. The questionnaire consisted of a set of problems, each of which could stand on its own, making its own point—I had long had the ambition of doing "psychology with single questions," and this was an opportunity to try it. Amos went off and administered the questionnaire to participants at a meeting of the Mathematical Psychology Association, and a few weeks later, we met in Jerusalem to look at the results and write a paper.

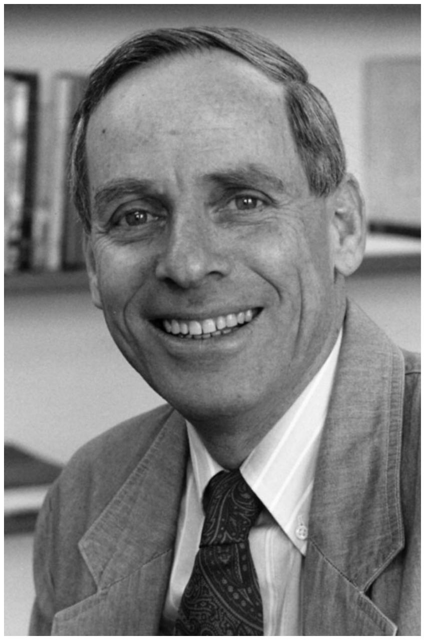
The experience was magical. I had enjoyed collaborative work before, but this was different. Amos was often described by people who knew him as the smartest person they knew. He was also very funny, with an endless supply of jokes appropriate to every nuance of a situation. In his presence, I became funny as well, and the result was that we could spend hours of solid work in continuous mirth. The paper we wrote was deliberately humorous—we described a prevalent belief in the "law of small numbers," according to which the law of large numbers extends to small numbers as well. Although we never wrote another humorous paper, we continued to find amusement in our work—I have probably shared more than half of the laughs of my life with Amos.

And we were not just having fun. I quickly discovered that Amos had a remedy for everything I found difficult about writing. With him, movement was always forward. Progress might be slow, but each successive draft we produced was an improvement—this was not something I could take for granted when working on my own. Amos's work was always characterized by confidence and by a crisp elegance, and it was a joy to find those characteristics now attached to my ideas as well. As we were writing our first paper, I was conscious of how much better it was than the more hesitant piece I would have written by myself. I

Editor's note. This excerpt was adapted with permission from a longer autobiographical statement to be published in *Les Prix Nobel 2002* (Frängsmyr, in press). Copyright by the Nobel Foundation. Adapted by permission.

Author's note. Correspondence concerning this article should be addressed to Daniel Kahneman, Department of Psychology, Princeton University, Princeton, NJ 08544-1010. E-mail: kahneman@princeton.edu

Vol. 58, No. 9, 723–730 DOI: 10.1037/0003-066X.58.9.723



Amos Tversky
Photo by Ed Souza, Stanford News Service

don't know exactly what it was that Amos found to like in our collaboration—we were not in the habit of trading compliments—but clearly he was also having a good time. We were a team, and we remained in that mode for well over a decade. The Nobel Prize was awarded for work that we produced during that period of intense collaboration.

At the beginning of our collaboration, we quickly established a rhythm that we maintained during all our years together. Amos was a night person, and I was a morning person. This made it natural for us to meet for lunch and a long afternoon together and still have time to do our separate things. We spent hours each day just talking. We did almost all the work on our joint projects while physically together, including the drafting of questionnaires and papers, and we avoided any explicit division of labor. Our principle was to discuss every disagreement until it had been resolved to our mutual satisfaction, and we had tie breaking rules for only two topics: whether an item should be included in the list of references (Amos had the casting vote) and who should resolve any issue of English grammar (my dominion). We did not initially have a concept of a senior author. We tossed a coin to determine the order of authorship of our first paper and alternated from then on until the pattern of our collaboration changed in the 1980s.

One consequence of this mode of work was that all our ideas were jointly owned. Our interactions were so frequent and so intense that there was never much point in distinguishing between the discussions that primed an idea, the act of uttering it, and the subsequent elaboration of it. I believe that many scholars have had the experience of discovering that they had expressed (sometimes even published) an idea long before they really understood its significance. It takes time to appreciate and develop a new thought. Some of the greatest joys of our collaboration—and probably much of its success—came from our ability to elaborate each other's nascent thoughts: If I expressed a half-formed idea, I knew that Amos would be there to understand it, probably more clearly than I did, and that if it had merit, he would see it. Like most people, I am somewhat cautious about exposing tentative thoughts to others—I must first make sure that they are not idiotic. In our years of close collaboration, this caution was completely absent. The mutual trust and the complete lack of defensiveness that we achieved were particularly remarkable because both of us—Amos even more than I—were known to be severe critics. Our magic worked only when we were by ourselves. We soon learned that joint collaboration with any third party should be avoided because we became competitive in a threesome.

Amos and I shared the wonder of together owning a goose that could lay golden eggs—a joint mind that was better than our separate minds. The statistical record confirms that our joint work was superior, or at least more influential, than the work we did individually (Laibson & Zeckhauser, 1998). Amos and I published eight journal articles during our peak years (1971–1981), of which five had been cited more than a thousand times by the end of 2002. Of our separate works, which in total number over 200, only Amos's theory of similarity (Tversky, 1977) and my book on attention (Kahneman, 1973) exceeded that threshold.

The special style of our collaborative work was recognized early by a referee of our first theoretical paper (on representativeness), who caused it to be rejected by *Psychological Review*. The eminent psychologist who wrote that review—his anonymity was betrayed years later—pointed out that he was familiar with the separate lines of work that Amos and I had been pursuing and considered both quite respectable. However, he added the unusual remark that we seemed to bring out the worst in each other and certainly should not collaborate. He found most objectionable our method of using multiple single questions as evidence—and he was quite wrong there as well.

The 1974 *Science* Article and the Rationality Debate

Amos and I were at the Oregon Research Institute (ORI) in Eugene from 1971 to 1972, a year that was by far the most productive of my life. We did a considerable amount of research and writing on the availability heuristic, on the psychology of prediction, and on the phenomena of anchoring and overconfidence—thereby fully earning the label *dynamic duo* that our colleagues attached to us. Working evenings and nights, I also completely rewrote my book *Attention and Effort* (Kahneman, 1973), which went to the publisher that year and remains my most significant independent contribution to psychology.

At ORI, I came into contact for the first time with an exciting community of researchers that Amos had known since his student days at the University of Michigan: Paul Slovic, Sarah Lichtenstein, and Robyn Dawes. I also learned much from Lewis Goldberg's work on clinical and actuarial judgment and from Paul Hoffman's ideas about paramorphic modeling. ORI was one of the major centers of judgment research, and I had the occasion to meet quite a few of the significant figures of the field when they came visiting, Ken Hammond among them.

Some time after our return from Eugene to Jerusalem, Amos and I settled down to review what we had learned about three heuristics of judgment (representativeness, availability, and anchoring) and about a list of a dozen biases associated with these heuristics. We spent a delightful year in which we did little but work on a single article. On our usual schedule of spending afternoons together, a day on which we advanced the paper by a sentence or two was considered quite productive. Our enjoyment of the process gave us unlimited patience. We wrote as if the precise choice of every word were a matter of great moment.

We published the article (Tversky & Kahneman, 1974) in *Science* because we thought that the prevalence of systematic biases in intuitive assessments and predictions could possibly be of interest to scholars outside psychology. The *Science* article turned out to be a rarity: an empirical psychological article that (some) philosophers and (a few) economists could and did take seriously. What was it that made readers of the article willing to listen? I attribute the unusual attention at least as much to the medium as to the message. Amos and I had continued to practice the psychology of single questions, and the *Science* article—like others we wrote—incorporated questions that were cited verbatim in the text. These questions, I believe, personally engaged the readers and convinced them that we were concerned not with the stupidity of Joe Public, but with a much more interesting issue: the susceptibility to erroneous intuitions of intelligent, sophisticated, and perceptive individuals such as themselves. Whatever the reason, the article soon became a standard reference as an attack on the rational-agent model, and it spawned a large literature in cognitive science, philosophy, and psychology. We had not anticipated that outcome.

I realized only recently how fortunate we were not to have aimed deliberately at the large target we happened to hit. If we had intended the article as a challenge to the rational model, we would have written it differently, and the challenge would have been less effective. An essay on rationality would have required a definition of that concept, a treatment of boundary conditions for the occurrence of biases, and a discussion of many other topics about which we had nothing interesting to say. The result would have been less crisp, less provocative, and ultimately less defensible. As it was, we offered a progress report on our study of judgment under uncertainty, which included much solid evidence. All inferences about human rationality were drawn by the readers themselves.

The conclusions that readers drew were often too strong, mostly because existential quantifiers, as they are prone to do, disappeared in the transmission. Whereas we had shown that (some, not all) judgments about uncertain events are mediated by heuristics, which (sometimes, not always) produce predictable biases, we were often read, both by friendly readers and by critics, as having claimed that people cannot think straight. Our allegedly negative view of the human mind was criticized by quite a few scholars (among them Cohen, 1981; Gigerenzer, 1991, 1996; and Lopes, 1991). The fact that men had walked on the moon was used more than once as an argument against our position. Because our treatment was mistakenly taken to be all-inclusive, our silences became significant. For example, the fact that we had written nothing about the role of social factors in judgment was taken as an indication that we thought these factors were unimportant. We thought nothing of the kind—all that happened was that we had nothing to say about social factors.

I now suspect that any body of work will eventually be misunderstood if it attracts attention long enough for the context in which it is read to be markedly different from the context in which it was first presented. Omissions are particularly likely to be misunderstood. In light of issues that were later raised about the robustness of cognitive illusions, for example, it may now appear odd that the early work did not discuss cases in which intuition is accurate or cases in which intuitive bias is replaced by correct reasoning. The following is a possible answer:

> The authors of the "law of small numbers" saw no need to examine correct statistical reasoning. They believed that including easy questions in the design would insult the participants and bore the readers. More generally, the early studies of heuristics and biases displayed little interest in the conditions under which intuitive reasoning is preempted or overridden—*controlled reasoning leading to correct answers was seen as a default case that needed no explaining* [italics added]. (Kahneman & Frederick, 2002, p. 50)

Another example is the role of affect in biases. In Kahneman (2003, this issue), I refer to the formulation of an *affect heuristic* by Paul Slovic (Slovic, Finucane, Peters, & MacGregor, 2002) as "probably the most important development in the study of judgment heuristics in the past few decades" (p. 710). After writing this sentence, I began to wonder about the obvious question: Why did Tversky and I completely neglect the role of affect in intuitive judgment? Much to my surprise, the answer was the same. When we began our work, the default interpretation of biases of judgment was that they were motivated, or otherwise driven by emotion. Indeed, one of the innovative features of our work was that we discussed biases that were not motivated or emotional!

In the newer context, which Tversky and I participated in creating, the ubiquity of correct reasoning and of affective biases no longer goes without saying. The failure to mention these effects therefore has a different meaning now than it did when we wrote our article. Current readers are bound to misunderstand our original intent because the conversational context has changed, and the inevitable misunderstanding of the work is due, at least in part, to the influence that it has had. I wonder how often that occurs.

The difficulties of communication had another side. Amos and I, of course, never forgot what we had thought about the quality of reasoning, and we naturally assumed that our meaning remained transparent (Keysar & Barr, 2002). From that vantage point, the critics who misinterpreted our early omissions as an attack on the human mind appeared to be willfully distorting the meaning of our work. This was not mere paranoia on our part: It is a fact of life that scholars who invest time and effort in the legitimate enterprise of criticism rarely adopt the most charitable interpretation of the works with which they disagree. However, the criticisms we saw as unfair were often directed at an interpretation that was also common among the friendly readers of our work, even if it was not the one we had intended.

I draw several lessons from my experience with the controversy about the quality of the human mind. The first is that, at least in the social sciences, any article that is more than 10 years old should always be read together with a brief description of the context of opinions in which it was written. The second is that scholars who plan to engage in an enterprise that is primarily critical should be encouraged to discuss the issues with their intended target, preferably over a drink, before they start processing sarcastic words. The third lesson is a more formal procedure for dealing with controversy, to which I return later.

Prospect Theory

After the publication of our article on judgment in *Science* (Tversky & Kahneman, 1974), Amos suggested that we study decision making together. This was a field in which he was already an established star and one about which I knew very little. For an introduction, he suggested that I read the relevant chapters of the text *Mathematical Psychology: An Elementary Introduction,* which he had coauthored (Coombs, Dawes, & Tversky, 1970). Utility theory and the paradoxes of Allais and Ellsberg were discussed in the book, along with some of the classic experiments in which major figures in the field had joined in an effort to measure the utility function for money by eliciting choices between simple gambles.

I learned from the book that the name of the game was the construction of a theory that would explain Allais's paradox parsimoniously. As psychological questions go, this was not a difficult one, because Allais's famous problems are, in effect, an elegant way to demonstrate that the subjective response to probability is not linear. The natural response of a decision theorist to the Allais paradox, certainly in 1975 and probably even today, would be to search for a new set of axioms that have normative appeal and yet permit the nonlinearity. The natural response of psychologists was to set aside the issue of rationality and to develop a descriptive theory of the preferences that people actually have, regardless of whether or not they should have them.

The task we set for ourselves was to account for observed preferences in the quaintly restricted universe in which the theory of choice has traditionally been studied: monetary gambles with few outcomes and definite probabilities. We were trying to answer an empirical question, and data were needed. Amos and I solved the data-collection problem with a method that was both efficient and pleasant. We spent our hours together inventing interesting choices and examining our preferences. If we agreed on the same choice, we provisionally assumed that other people would also accept it, and we went on to explore its theoretical implications. This unusual method enabled us to move quickly, and we constructed and discarded models at a dizzying rate. I have a distinct memory of a model that was numbered 37 but cannot vouch for the accuracy of our count.

As was the case in our work on judgment, our central insights were acquired early, and as in our work on judgment, we spent a vast amount of time and effort before publishing a paper that summarized those insights (Kahneman & Tversky, 1979). The first insight came as a result of my naïveté. When reading Coombs et al.'s (1970) mathematical psychology textbook, I was puzzled that all the choice problems were described in terms of gains and losses (actually, almost always gains), whereas the utility functions that were supposed to explain the choices were drawn with wealth as the abscissa. This seemed unnatural and psychologically unlikely. We immediately decided to adopt changes and/or differences as carriers of utility. We had no inkling that this obvious move was truly fundamental or that it would open the path to behavioral economics. Harry Markowitz, who won the Nobel Prize in economics in 1990, had proposed changes of wealth as carriers of utility in 1952, but he did not pursue the idea deeply, and it was not adopted by others in the field.

The shift from wealth to changes of wealth as carriers of utility is significant because of a property of preferences that we later labeled *loss aversion:* The response to losses is consistently much more intense than the response to corresponding gains, with a sharp kink in the value function at the reference point. Loss aversion is manifest in the extraordinary reluctance to accept risk that is observed when people are offered a gamble on the toss of a coin: Most will reject a gamble in which they might lose $20, unless they are offered more than $40 if they win. The concept of loss aversion was, I believe, our most useful contribution to the study of decision making. The asymmetry between gains and losses solves quite a few puzzles, including the widely noted and economically irrational distinction that people draw between opportunity costs and "real" losses. Loss aversion also helps explain why real-estate markets dry up for long periods when prices are down, and it contributes to the explanation of a widespread bias favoring the status quo in decision making. Finally, the asymmetric consideration of gains and losses extends to the domain of moral intuitions, in which imposing losses and failing to share gains are evaluated quite differently. But, of course, none of that was visible to Amos and me when we first decided to assume a kinked value function—we needed that kink to account for choices between gambles.

Another set of early insights came when Amos suggested that we flip the signs of outcomes in the problems we had been considering. The result was exciting. We immediately detected a remarkable pattern, which we called *reflection:* Changing the signs of all outcomes in a pair of gambles almost always caused the preference to change from risk averse to risk seeking or vice versa. For example, we both preferred a sure gain of $900 over a .9 probability of gaining $1,000 (or nothing), but we preferred a gamble with a .9 probability of losing $1,000 over a sure loss of $900. We were not the first to observe this pattern. Raiffa (1968) and Williams (1966) knew about the prevalence of risk seeking in the negative domain, but ours was apparently the first serious attempt to make something of it.

We soon had a draft of a theory of risky choice, which we called *value theory* and presented at a conference in the spring of 1975. We then spent about three years polishing it, until we were ready to submit the article for publication. Our effort during those years was divided between the tasks of exploring interesting implications of our theoretical formulation and developing answers to all plausible objections. To amuse ourselves, we invented the specter of an ambitious graduate student looking for flaws, and we labored to make that student's task as thankless as possible. The most novel idea of value theory (later named *prospect theory*) occurred to us in that defensive context. It came quite late, as we were preparing the final version of the paper. We were concerned that a straightforward application of our model implied that the prospect denoted ($100, .01; $100, .01)—two mutually exclusive .01 chances to gain $100—is more valuable than the prospect ($100, .02). The prediction is wrong, of course, because most decision makers will spontaneously transform the former prospect into the latter and treat them as equivalent in subsequent operations of evaluation and choice. To eliminate the problem, we proposed that decision makers, prior to evaluating the prospects, perform an editing operation that collects similar outcomes and adds their probabilities. We went on to propose several other editing operations that provided an explicit and psychologically plausible defense against a variety of superficial counterexamples to the core of the theory. We had succeeded in making life quite difficult for that pedantic graduate student, but we had also made a truly significant advance, by making it explicit that the objects of choice are mental representations, not objective states of the world. This was a large step toward the development of a concept of framing and eventually toward a new critique of the model of the rational agent.

When we were ready to submit the work for publication, we deliberately chose a meaningless name for our theory: *prospect theory*. We reasoned that if the theory ever became well known, having a distinctive label would be an advantage. We published the article (Kahneman & Tversky, 1979) in *Econometrica*. The choice of venue turned out to be important; the identical article published in *Psychological Review* would likely have had little impact on economics. But our decision was not guided by a wish to influence economics. *Econometrica* just happened to be the journal where the best articles on decision making had been published, and we were aspiring to be in that company.

In another way, the impact of prospect theory depended crucially on the medium, as well as the message. Prospect theory was a formal mathematical theory, and its formal nature was the key to the impact it had in economics. Every discipline of social science, I believe, has some ritual tests of competence that must be passed before a piece of work is considered worthy of attention. Such tests are necessary to prevent information overload, and they are also important aspects of the tribal life of the disciplines. In particular, they allow insiders to ignore just about anything that is done by members of other tribes and to feel no scholarly guilt about doing so. To serve this screening function efficiently, the competence tests usually focus on some aspect of form or method and have little or nothing to do with substance. Prospect theory passed such a test in economics, and its observations became a legitimate (though optional) part of the scholarly discourse in that discipline. It is a strange and rather arbitrary process that selects some pieces of scientific writing for relatively enduring fame while committing most of what is published to almost immediate oblivion.

Framing and Mental Accounting

Amos and I completed prospect theory during the 1977–1978 academic year, which I spent at the Center for Advanced Studies for the Social and Behavioral Sciences and Amos spent as a visitor at the psychology department of Stanford University. Around that time, we began work on our next project, which became the study of framing. This was also the year in which the second most important professional friendship in my life—with Richard Thaler—had its start.

A framing effect is demonstrated by constructing two transparently equivalent versions of a given problem, which nevertheless yield predictably different choices. The standard example of a framing problem, which was developed quite early, is the "lives saved, lives lost" question, which offers a choice between two public-health programs proposed to deal with an epidemic that is threatening 600 lives: One program will save 200 lives; the other has a one-third chance of saving all 600 lives and a two-thirds chance of saving none. In this version, people prefer the program that will save 200 lives for sure. In the second version, one program will result in 400 deaths, and the other has a two-thirds chance of 600 deaths and a one-third chance of no deaths. In this formulation, most people prefer the gamble. If the same respondents are given the two problems on separate occasions, many give incompatible responses. When confronted with their inconsistency, people are quite embarrassed and usually lose confidence in both their previous answers.

Amos and I began creating pairs of problems that revealed framing effects while we were working on prospect theory. We used them to show sensitivity to gains and losses (as in the lives example) and to illustrate the inadequacy of a formulation in which the only relevant outcomes are final states. In that article, we also showed that a single-stage gamble could be rearranged as a two-stage gamble in a manner that left the bottom-line probabilities and outcomes unchanged but reversed preferences. Later, we developed choice problems that could be presented either as a single choice or as a sequence of two choices, leading to inconsistent preferences and to preferences for strictly inferior options.

These are not parlor-game demonstrations of human stupidity. The ease with which framing effects can be demonstrated reveals a fundamental limitation of the human mind. Framing effects violate a basic requirement of rationality, one that we called *invariance* (Kahneman & Tversky, 1984) and Arrow (1982) called *extensionality*. It took us a long time and several iterations to develop a forceful statement of this contribution to the rationality debate, which we presented several years after our framing article (Tversky & Kahneman, 1986).

Another advance that we made in our first framing article was the inclusion of riskless choice problems among our demonstrations of framing. In making that move, we had help from a new friend. Richard Thaler was a young economist blessed with a sharp and irreverent mind. While still in graduate school, he had trained his ironic eye on his own discipline and had collected a set of pithy anecdotes demonstrating obvious failures of basic tenets of economic theory in the behavior of people in general—and of his very conservative professors in Rochester in particular. One key observation was the endowment effect, which Dick illustrated with the example of the owner of a bottle of old wine, who would refuse to sell it for $200 but would not pay as much as $100 to replace it if it broke. Sometime in 1976, a copy of the 1975 draft of prospect theory got into Dick's hands, and that event made a significant difference to our lives. Dick realized that the endowment effect, which is a genuine puzzle in the context of standard economic theory, is readily explained by two assumptions derived from prospect theory. First, the carriers of utility are not states (owning or not owning the wine), but changes—getting the wine or giving it up. And because of loss aversion, giving up is weighted more than getting. When Dick learned that Amos and I would be in Stanford, he secured a visiting appointment at the Stanford branch of the National Bureau of Economic Research, which is located on the same hill as the Center for Advanced Studies. We soon became friends and have ever since had a considerable influence on each other's thinking.

The endowment effect was not the only thing we learned from Dick. He had also developed a list of phenomena of what we now call *mental accounting*. The concept of mental accounting describes how people violate rationality by failing to maintain a comprehensive view of outcomes and by failing to treat money as fungible. Dick showed how people segregate their decisions into separate accounts, then struggle to keep each of these accounts in the black. He inspired me to invent another story, in which a person who comes to the theater realizes that he has lost his ticket (in one version) or has lost an amount of cash equal to the ticket value (in another version). People report that they would be very likely still to buy a ticket if they had lost the cash, presumably because the loss has been charged to general revenue. On the other hand, they describe themselves as quite likely to go home if they have lost an already purchased ticket, presumably because they do not want to pay twice to see the same show.

Behavioral Economics

Our interaction with Thaler eventually proved to be more fruitful than we could have imagined at the time, and it was a major factor in the awarding of the Nobel Prize. The committee cited me "for having integrated insights from psychological research into economic science" (*Press Release,* 2002). Although I do not wish to renounce any credit for my contribution, I should say that in my view, the work of integration was actually done mostly by Thaler and the group of young economists that quickly began to form around him, starting with Colin Camerer and George Loewenstein and followed by Matthew Rabin, David Laibson, Terry Odean, Sendhil Mullainathan, and Nick Barberis. Amos and I provided quite a few of the initial ideas that were eventually integrated into the thinking of these economists, and prospect theory afforded legitimacy to the enterprise of drawing on psychology as a source of realistic assumptions about economic agents. But the founding text of behavioral economics was the first article in which Thaler (1980) presented a series of vignettes that challenged fundamental tenets of consumer theory. The respectability that behavioral economics now enjoys within the discipline was secured, I believe, by some important discoveries Dick made in what is now called behavioral finance and by the series of "Anomalies" columns that he published in every issue of the *Journal of Economic Perspectives* from 1987 to 1990 and has continued to write occasionally since that time.

In 1982, Amos and I attended a meeting of the Cognitive Science Society in Rochester, where we had a drink with Eric Wanner, a psychologist who was then vice president of the Sloan Foundation. Eric told us that he was interested in promoting the integration of psychology and economics and asked for our advice on ways to go about it. I have a clear memory of the answer we gave him. We thought that there was no way to "spend a lot of money honestly" on such a project because interest in interdisciplinary work could not be coerced. We also thought that it was pointless to encourage psychologists to make themselves heard by economists but that it could be useful to encourage and support the few economists who were interested in listening. Thaler's name surely came up. Soon after that conversation, Wanner became the president of the Russell Sage Foundation, and he brought the psychology–economics project with him. The first grant that he made in that program was for Dick Thaler to spend an academic year (1984–1985) visiting me at the University of British Columbia in Vancouver.

That year was one of the best in my career. We worked as a trio that also included the economist Jack Knetsch, with whom I had already started constructing surveys on a variety of issues, including valuation of the environment and public views about fairness in the marketplace. We did a lot together that year. We conducted a series of market experiments involving real goods (the "mugs studies"), which eventually became a standard in that literature (Kahneman, Knetsch, & Thaler, 1990). We also conducted multiple surveys in which we used experimentally varied vignettes to identify the rules of fairness that the public would apply to merchants, landlords, and employers (Kahneman, Knetsch, & Thaler, 1986b). Our central observation was that in many contexts, the existing situation (e.g., price, rent, or wage) defines a *reference transaction* to which the transactor (consumer, tenant, or employee) has an entitlement—the violation of such entitlements is considered unfair and may evoke retaliation. For example, cutting the wages of an employee merely because he could be replaced by someone who would accept a lower wage is unfair, although paying a lower wage to the replacement of an employee who quit is entirely acceptable. We submitted the paper to the *American Economic Review* and were utterly surprised by the outcome: The paper was accepted without revision. Luckily for us, the editor had asked two economists quite open to our approach to review the paper. We eventually learned that one of them was George Akerlof, who later became a friend.

Questions that arose during this research included whether people would be willing to pay something to punish another agent who they felt had treated them unfairly and whether people would in some circumstances share a windfall with a stranger in an effort to be fair. We decided to investigate these ideas using experiments for real stakes. The games that we invented for this purpose have become known as the ultimatum game and the dictator game. Alas, while writing up our second paper on fairness (Kahneman, Knetsch, & Thaler, 1986a), we learned that we had been scooped on the ultimatum game by Werner Güth and his colleagues, who had published experiments using the same design a few years earlier. I remember being quite crestfallen when I learned this. I would have been much more depressed if I had known how important the ultimatum game would eventually become.

That was the closest I ever came to doing economics. Since that time, I have been involved in other topics of research, and I have been cheering Thaler and behavioral economics from the sidelines. There has been much to cheer about. As a mark of the progress that has been made, I recall that when Matthew Rabin joined the Berkeley economics department as a young assistant professor and chose to immerse himself in psychology, many considered the move professional suicide. Some 15 years later, Rabin has earned the prestigious Clark Medal for the best economist under age 40, and George Akerlof (2002) has delivered a Nobel lecture on the topic of behavioral macroeconomics.

Eric Wanner and the Russell Sage Foundation have continued to support behavioral economics over the years. I was instrumental in the idea of using some of that support to set up a summer school for graduate students and young faculty in that field, and I helped Dick Thaler and Colin Camerer organize the first one in 1994. When the fifth summer school convened in 2002, David Laibson, who had been a participant in 1994, was tenured at Harvard and was one of the three organizers. Terrance Odean and Sendhil Mullainathan, who had also participated as students, came back to lecture as successful researchers with positions in two of the best universities in the world. It was a remarkable experience to hear Matthew Rabin teach a set of guidelines for developing theories in behavioral economics—including the suggestion that the standard economic model should be a special case of the more complex and general models that were to be constructed. We had come a long way.

Although behavioral economics has enjoyed much more rapid progress and gained more respectability in economics than appeared possible 15 years ago, it is still a minority approach, and its influence on most fields of economics is negligible. Many economists believe that it is a passing fad, and some hope that it will be. The future may prove them right. But many bright young economists are now betting their careers on the expectation that the current trend will last for some time. And such expectations have a way of being self-fulfilling.

Adversarial Collaboration

One of the lessons I have learned from a long career is that controversy is a waste of effort. I take some pride in the fact that there is not one item in my bibliography that was written as an attack on someone else's work, and I am convinced that the time I spent on a few occasions in reply–rejoinder exercises would have been better spent doing something else. Both as a participant and as a reader, I have been appalled by the absurdly competitive and adversarial nature of these exchanges, in which hardly anyone ever admits an error or acknowledges learning anything from the other. Doing angry science is a demeaning experience—I have always felt diminished by the sense of losing my objectivity when in point-scoring mode.

Some years ago, I decided to do something about this. Unaware that someone else had preceded me (Latham, Erez, & Locke, 1988), I began to champion a procedure of *adversarial collaboration* as a substitute for the format of critique–reply–rejoinder in which debates are currently conducted in the social sciences. Adversarial collaboration involves a good-faith effort to conduct debates by carrying out joint research—in some cases an agreed-upon arbiter may be needed to lead the project and collect the data. Because the contestants are not expected to reach complete agreement at the end of the exercise, adversarial collaborations will usually lead to an unusual type of joint publication, in which disagreements are laid out as part of a jointly authored paper. I have had three adversarial collaborations, with Tom Gilovich and Victoria Medvec (Gilovich, Medvec, & Kahneman, 1998), with Ralph Hertwig (where Barbara Mellers was the agreed-upon arbiter; see Mellers, Hertwig, & Kahneman, 2001), and with a group of experimental economists in the United Kingdom (Bateman, Kahneman, Munro, Starmer, & Sugden, 2003). All three ended with some new facts accepted by all, narrowed differences of opinion, and considerable mutual respect. An appendix in Mellers et al.'s (2001) article proposed a detailed protocol for the conduct of adversarial collaboration.

In another case, I did not succeed in convincing two colleagues that we should engage in an adversarial collaboration, but we jointly developed another procedure that is also more constructive than the reply–rejoinder format. They wrote a critique of one of my lines of work, but instead of following up with the usual exchange of unpleasant comments, we decided to write a joint piece, which started with a statement of what we did agree on, then went on to a series of short debates about issues on which we disagreed (Ariely, Kahneman, & Loewenstein, 2000). The result was much more enlightening than what we would have done in the conventional format.

My hope is that these and other variants of adversarial collaboration may eventually become standard. This is not a mere fantasy: It would be easy for journal editors to require critics of the published work of others—and the targets of such critiques—to make a good-faith effort to explore differences constructively. I believe that the establishment of such procedures would contribute to an enterprise that more closely approximates the ideal of science as a cumulative social product.

REFERENCES

Akerlof, G. A. (2002). Behavioral macroeconomics and macroeconomic behavior. *American Economic Review, 92,* 411–433.

Ariely, D., Kahneman, D., & Loewenstein, G. (2000). Joint comment on "When does duration matter in judgment and decision making?" *Journal of Experimental Psychology: General, 129,* 524–529.

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

Bateman, I. J., Kahneman, D., Munro, A., Starmer, C., & Sugden, R. (2003). *Is there loss aversion in buying? An adversarial collaboration to discriminate between alternative theories of loss aversion.* Manuscript in preparation, University of East Anglia, Norwich, England.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317–331.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction.* Oxford, England: Prentice-Hall.

Frängsmyr, T. (Ed.). (in press). *Les Prix Nobel 2002* [Nobel Prizes 2002]. Stockholm, Sweden: Almqvist & Wiksell International.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2., pp. 83–115). Chichester, England: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103,* 592–596.

Gilovich, T., Medvec, V. H., & Kahneman, D. (1998). Varieties of regret: A debate and partial resolution. *Psychological Review, 105,* 602–605.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986a). Fairness and the assumptions of economics. *Journal of Business, 59,* S285–S300.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986b). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76,* 728–741.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98,* 1325–1348.

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

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they "should." In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 150–166). New York: Cambridge University Press.

Laibson, D., & Zeckhauser, R. (1998). Amos Tversky and the ascent of behavioral economics. *Journal of Risk and Uncertainty, 16,* 7–47.

Latham, G. P., Erez, M., & Locke, E. A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting. *Journal of Applied Psychology, 73,* 753–772.

Lopes, L. A. (1991). The rhetoric of irrationality. *Theory and Psychology, 1,* 65–82.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Press release: The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2002. (2002, October 9). Retrieved September 10, 2003, from http://www.nobel.se/economics/laureates/2002/press.html

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison-Wesley.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 397–420). New York: Cambridge University Press.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1,* 39–60.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

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

Williams, A. C. (1966). Attitudes toward speculative risks as an indicator of attitudes toward pure risks. *Journal of Risk and Insurance, 33,* 577–586.