*Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)*
*http://dare.uva.nl/document/130816*

| | |
|---|---|
| File ID | 130816 |
| Filename | 3: Tversky's behavioral deviations and Kahneman's cognitive mistakes |
| Version | Final published version (publisher's pdf) |

SOURCE (OR PART OF THE FOLLOWING SOURCE):

| | |
|---|---|
| Type | Dissertation |
| Title | Kahneman and Tversky and the making of behavioral economics |
| Author | F. Heukelom |
| Faculty | Faculty of Economics and Business |
| Year | 2009 |
| Pages | V, 177 |
| ISBN | 978 90 361 0125 7 |

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/301297

*Copyright*

*It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other then for strictly personal, individual use.*

*UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)*

## 3. Tversky's behavioral deviations and Kahneman's cognitive mistakes

### 1. Kahneman and Tversky before Kahneman and Tversky

The previous chapter showed how in postwar American psychology measurement theory was central to the development of mathematical psychology. This was further reflected in the newly-created field of behavioral decision research in psychology that assigned itself the task of testing decision theory's axioms of normative behavior experimentally. This was done with the purpose of engineering solutions for cases in which the human decision making machinery failed. Because mathematical psychology's measurement theory used the human being as a measurement instrument, and because behavioral decision research required a measurement instrument to conduct its experiments, mathematical psychology and behavioral decision research were natural extensions of one another. What was important for their interrelationship was furthermore that both used the University of Michigan as their principal base, although contributors to both programs also came from other universities.

In 1965, Amos Tversky (1937-1996) obtained his PhD from the University of Michigan under the supervision of Clyde Coombs and Ward Edwards. Tversky embodied the close connection between mathematical psychology and behavioral decision research. His dissertation combined Edwards' focus on normative rules for rational behavior with Coombs' interest in descriptive theories of measurement. Within a few years after finishing his PhD, Tversky became one of the foremost contributors to mathematical psychology's representational theory of measurement and was one of the four authors of *Foundations of Measurement*, the standard work in the representational theory of measurement. Besides his work in mathematical psychology Tversky collaborated with Edwards and quickly became a frontrunner in behavioral decision research. Tversky's prominence in behavioral decision research is exemplified by the fact that the collection of core behavioral decision research publications, *Decision Making* (1967), was co-edited by Edwards and Tversky.

Tversky's greatest claim to fame came in the 1970s from his work with Daniel Kahneman (1934- ). Together Kahneman and Tversky constructed a new approach to human decision behavior in behavioral decision research that soon became more prominent than the approach favored by Edwards. They termed the new approach 'Heuristics and Biases.' Heuristics and Biases and its derivative prospect theory laid

the basis for behavioral economics which will be dealt with in Chapter four. Before we look at Kahneman and Tversky's collaborative research in the 1970s, however, we need to understand why Tversky became dissatisfied with the framework of behavioral decision research in the 1960s. To understand Kahneman and Tversky's approach to behavioral decision research of the 1970s, it is necessary to understand the problems that Tversky tried to solve. The first section of this chapter thus deals with Tversky's work in mathematical psychology and behavioral decision research in the 1960s.

But Kahneman and Tversky's work of the 1970s was not solely the product of Tversky's mind. Indeed, it is my belief that the crucial twist that solved Tversky's problems in behavioral decision research came from Kahneman. An experimental psychologist just as Tversky, Kahneman came from a different theoretical background. The research that Kahneman conducted during the 1960s included the psychophysics of vision, semantic differentials, trade-offs between different cognitive tasks and related issues. The recurring theme in Kahneman's work in the 1960s, and in his whole career, is that of cognitive mistakes. In all his research Kahneman investigates when human beings make cognitive mistakes, how severe these mistakes are, what explanation might account for this mistaken decision making, and in what way can mistakes be prevented in the future. In Kahneman's work experimental psychology's emphasis on deviations from the norm behavior became strongly pronounced. To understand why and in what way Kahneman altered Tversky's behavioral decision research, we therefore need to go back to Kahneman's work in the 1960s. The second section of this chapter deals with Kahneman's work in the 1960s.

## 2. Caught between *a priori* axioms and behavioral deviations: Tversky's research in the 1960s

Between finishing his PhD in 1965 and completing his first publication with Kahneman in 1971 Tversky worked simultaneously in three related areas. He worked within Leonard Savage's decision theory on the further refinement of normative decision theory, he extensively contributed to the mathematical development of the representational theory of measurement in mathematical psychology, and he measured the actual decision behavior of human beings in experiments and compared these measurements to the norms of decision theory. In the section which follows, I will

discuss Tversky's work in these three areas separately, before making an overall assessment of Tversky's work during the first six years following his PhD dissertation.

2.1 *Decision theory*

In his work on decision theory in the 1960s, Tversky adhered to the model and approach set out by Savage: "Utility theory, or the subjective expected utility (SEU) model, is a theory about decision making under risk. It is based on a set of seemingly reasonable axioms (Savage, 1954) which imply that an individual's choices between risky alternatives can be described as the maximization of his subjective expected utility" [Tversky (1967c), p.27]. Tversky also accepted Edwards' interpretation and application of Savage's decision theory in terms of psychologists' normative-descriptive distinction. Thus, we read that utility theory "has been widely applied as a normative principle in economics and operations research, it underlies game theory and detection theory, and it has stimulated extensive experimental investigation" [Tversky (1967c), p.27].

For instance, in Tversky's first article "On the Optimal Number of Alternatives at a Choice Point" (1964), he described a mathematical model that determines the optimal number of alternatives at a choice point in a test. It considered as examples choices in "[m]ultiple choice tests, mazes or personality checks" [Tversky (1964), p.386], but the argument made implied that the theory could be applied to any choice problem in tests. The paper's model showed mathematically that the optimal number of alternatives at a choice point in terms of discriminability, power and information is three. Tversky applied the approach of decision theory to a specific problem that as yet had not been solved satisfactorily, and he subsequently constructed a norm for this specific problem. Applying decision theory to the number of options at a choice point in a test yielded an optimum of three. A rational professor should thus give his students multiple choice tests with always three alternatives.

For many decision situations it had already been established how a rational individual should behave. But often the mathematics of the solution to these decision situations could be improved. The theory was there, but it needed to be worked out a bit more. In the brief theoretical part of Tversky and Russo (1969), for example, the authors investigated the fundamental principle underlying "probabilistic theories of choice behavior" [Tversky and Russo (1969), p.1]. This fundamental principle, the

authors argued, appeared in different parts of the literature in different forms. It was known as the assumption of simple scalability, strong stochastic transitivity, substitutability, or independence. The authors showed that when these four principles were formulated in mathematical terms they amounted to the same. Tversky and Russo showed how what appeared on the surface to be different types of psychological investigations, in fact share a similar mathematical structure.

Tversky's work on decision theory in the 1960s accepted the decision theoretical work of Savage in particular as a starting point and refined it by applying it to specific situations and linking it to other theories. In his work on decision theory, Tversky was an exponent of the mathematical psychology in which he obtained his PhD. Tversky's purpose was not to come up with a new theory or to criticize a theory. Instead, the focus was on refining mathematically what was already there.

### 2.2 *Measurement theory*

A major part of Tversky's research during the period 1964-1971 was on measurement theory. Five lengthy articles in this period discuss topics such as the development of a generalized model of conjoint measurement, which allows one to measure probabilities and utilities in one and the same experiment; the foundations of multidimensional scaling; and multidimensional representation, which looks for ways to decompose different dimensions of measured dissimilarity judgments.[20] The main question of this research, as set out in Chapter two, was to investigate which mathematical structure the measurement procedure requires so as to measure what it should measure. An important issue was how the experiment should be conducted so that both the subjective probability and the subjective value could be derived from one and the same choice of the subjects in the experiments.

In this measurement literature Tversky equated measurement theory with decision theory. Tversky was a prominent contributor to measurement theory in mathematical psychology and his views, in this regard, were the same as the mathematical psychology community at large, as depicted in Chapter two. To Tversky, the individual in psychological experiments served as the measurement

[20] In the preface of *Foundations of Measurement* (1971), Krantz, Luce, Suppes, and Tversky acknowledge that all the articles on measurement that appeared in the period 1967-1970 by one, or a combination of the authors, were by-products of work on the book. That is, what is in these articles can also be found in the book. The five articles on measurement written by Tversky before 1971, four of which are co-publications with Krantz, were all written during these four years.

instrument, just like the thermometer or pan balance in the experiments of the physicist and chemist. As a consequence, the axioms of the representational theory of measurement that described the working of the measurement instrument were exactly the same as the axioms that described rational, normative behavior in decision theory. As this is a crucial part of Tversky's psychology, I will quote again from *Foundations of Measurement*.

> Unlike most theories of measurement, which may have both physical and behavioral interpretations, the theory of expected utility is devoted explicitly to the problem of making decisions when their consequences are uncertain. It is probably the most familiar example of a theory of measurement in the social sciences. [Krantz et al. (1971), p.369]

To Tversky, the axioms of normative decision theory were at the same time axioms that described the functioning of the psychologist's measurement instrument and axioms that described optimal decision behavior. Conceptually, this is only possible when one assumes that actual human decision behavior deviates very little from the norms of decision behavior. It assumes that if actual decision behavior deviates from the norm, it is somehow distributed evenly around the norm and does not deviate from the norm systematically. It also assumes that if human beings are found to deviate from the norms in certain situations, their mistakes can be relatively easily explained to them after which the individual will correct his or her behavior. Thus, when Tversky developed a conjoint measurement model for the representational theory of measurement using one measurement instrument to measure both utilities and probabilities, he needed to know how the subjects/instrument would behave in the experiment. For Tversky, the axioms of the measurement theory and of decision theory would predict how the subject/instrument would behave. On the basis of this knowledge, Tversky could then devise a measurement model and experiment that would produce both the perceived utilities and the perceived probabilities.

### 2.3 *Behavioral decision research*

Tversky's work on the representational theory of measurement and on decision theory came together in his experimental work on behavioral decision research. The measurement models were applied in experimental testing and actual behavior was

tested against normative decision theory. It is important to emphasize that in this early experimental work of Tversky, the question was *not* whether, let alone how, human beings deviate from the norms of decision theory. Tversky's basic research question in the 1960s, just as that posed by Edwards, was how to apply the normative model to human decision making behavior. The axioms of decision theory indicated how we should, and usually do make decisions given the utilities and probabilities implied by the different options. But what the axioms neglected to specify was how a human being perceives utilities and probabilities. In Savage's subjective expected utility (SEU) model, for instance, it was assumed that subjects have a subjective perception of both value and probabilities, termed utility and subjective probability respectively, which differs from the objective values and that subjects base their decisions on these subjective values. In a closely related model, the subjective expected value (SEV) model, it was assumed that subjects have only a subjective perception of the probabilities. Furthermore, it was investigated whether the subjective value curve of different goods has the same shape. It could, for instance, be the case that the subjective perception of the utility of candy decreases much faster than that of cigarettes. Thus, Tversky's experiments were first and foremost an investigation of how to apply the axioms of measurement theory and decision theory. The human being was used as a measurement instrument to measure the different attributes of actual human decision behavior, which in turn informed Tversky how decision theory best fitted into the experimental psychological framework.

At the same time, Tversky used his experiments to test whether the human instrument indeed functions properly. Tversky's experiments were set up to measure the subjective value curve of candy and cigarettes, but the experiments at the same time checked whether the subjects behaved according to the axioms of decision theory and measurement theory. As, for instance, the axiom of transitivity was the "cornerstone of normative and descriptive theories," and underlied "measurement models of sensation and value" [Tversky (1969), p.31], the experiments were used as an opportunity to also test the axioms of measurement theory. Thus, in one and the same experiment Tversky would apply the representational theory of measurement, check whether its measurement instrument functioned properly, measure subjective perception of probabilities and utilities, and monitor whether human beings indeed behave according to the axioms of decision theory.

For example, Tversky (1967a) tested the additivity and the independence axiom, two key axioms of normative decision theory and measurement theory, in a gambling experiment with eleven male inmates at the State Prison of Southern Michigan. The subjects had to gamble for, and were paid in, candy and cigarettes. Six normative decision models were compared for both the set of candy gambles and the set of cigarettes gambles. In both cases, it turned out, Savage's normative SEU model provided the best description of the behavior displayed by the subjects. Given the SEU model, both additivity and the independence of subjective probability and utility were confirmed. That is, assuming that people make their decisions according to the normative theory, the SEU model provided the best description. Nevertheless, Tversky was cautious and concluded that "After more than 15 years of experimental investigation of decisions under risk, the evidence on the descriptive validity of the SEU model is still inconclusive" [Tversky (1967a), p.199].

In a follow-up paper, Tversky (1967b) set out a measurement model that tested the descriptive validity of different normative models of decision making, among them Savage's SEU model, the power utility theory, and the strict additive model. To do so, the eleven inmates from the State Prison of Southern Michigan had to choose between different gambles, but did not know the relevant probabilities beforehand. Savage's SEU model provided the best description, but failed in the sense that the subjects consistently overestimated low and underestimated high probabilities. The model was therefore extended with a power utility function that allowed utility to be a non-monotonic function of money (versus monotonic in the standard case), and to vary across individuals. Tversky stressed the proven independence of subjective probability and utility in the experiment and concluded that the best descriptive model (SEU plus power utility) was incompatible with utility theory. In other words, to maintain the assumption that subjects make their decisions in the normatively correct way, Tversky had to assume a normative model that is inconsistent with utility theory. To maintain the idea that individuals make decisions rationally, it had to be concluded that utility theory is descriptively wrong.

Thus Tversky went a step further than Edwards in testing the axioms of measurement theory and decision theory. For Edwards the axioms were *a priori* truths that could not be violated. If the experimental results indicated that the axioms had

been systematically violated, then there must have been a problem with the experimental design somewhere. During the period between 1965 and 1971, however, Tversky came to believe that the axioms were systematically violated by the experimental subjects. In his experimental work, Tversky consistently distinguished between the normative and the descriptive. The emphasis was on well-known decision problems such as making gambles, but other decision situations that were investigated were, for instance, whether people can determine which of two lights is the brighter, or from which of two distributions a four-digit number can be drawn. Without exception, the experiments tested one or more normative models descriptively. That is, the hypothesis was that the normative model was a good description of actual decision behavior, which was then tested experimentally.

In a collaborative paper Tversky and Edwards (1966) investigated whether subjects seek the optimal amount of information that normative decision theory predicts that they will seek. To test this, an experiment was conducted in which subjects had to determine which of two lights is brighter. Subjects could obtain information about their results by paying with the money they had earned by giving correct responses. In some of the experimental treatments the subjects were told beforehand the distribution of each of the two lights. The normative model that was proposed as a descriptive model did not work well; subjects sought much more information than the model had predicted they would. The normative model which served as description of actual choice behavior was subsequently rejected. But the authors did not really know what to conclude from these results. They did not draw the conclusion that the normative model had been falsified, but they provided a number of explanations that might account for the deviations. At the same time, however, it was stressed that these explanations only partially explained the deviations. They concluded that for reasons yet to be discovered the normative model did not work well in this particular situation.

The problem of systematic deviations continued to bother Tversky and in 1969 he published a paper, entitled "Intransitivity of Preferences," in which experiments were described and discussed that had the sole purpose of testing the axiom of transitivity. Transitivity, Tversky stated, "is of central importance to both psychology and economics. It is the cornerstone of normative and descriptive decision theories."

Furthermore, it is the essential assumption in measurement theories since "it is a necessary condition for the existence of an ordinal (utility) scale" [Tversky (1969), p.31]. The article described a number of experiments that falsify weak stochastic transitivity (WST). In WST, transitivity of preferences is defined in terms of probabilities, hence, x is weakly preferred over y if and only if $P(x,y) \geq ½$, meaning that the probability of choosing x over y is larger than or equal to a half. It was shown that WST does not hold descriptively. That is to say, the subjects' actual decisions were not in the least stochastically transitive. And as transitivity is a key assumption in decision theory and measurement theory, this was potentially a serious problem as it implied that no normative model whatsoever could describe subjects' actual decision making.

But Tversky was still reluctant to give up on this foundation of both measurement theory and decision theory, as transitivity "is one of the basic and the most compelling principles of rational behavior" [Tversky (1969), p.45]. Tversky suggested that normative decision theory could be maintained because apparent intransitivities could always be attributed to an unobserved change of preference that takes place between the decisions made. He concluded somewhat paradoxically that "The main interest in the present results lies not so much in the fact that transitivity can be violated but rather in what these violations reveal about the choice mechanism and the approximation method that govern preference between multidimensional alternatives" [Tversky (1969), p.46].

The reason that Tversky was reluctant to accept that the experiments falsified the axioms was that the axioms were such a fundamental aspect of both measurement theory and decision theory. If he had accepted that the axioms were wrong, the representational theory of measurement and normative decision theory as description of human behavior would be falsified. Another reason why Tversky was reluctant to give up the axioms was that the experimental results did not always indicate falsification. In the above-mentioned experiments conducted in the prison of Southern Michigan, the decision behavior of the subjects largely corresponded to the norm. There were more situations in which the results were mixed. In Rapoport and Tversky (1970), subjects were presented with a sequence of offers and they had to decide at each stage whether to stop and take the present offer, or to continue sampling more

offers. The optimal stopping point was given by the normative analysis. It can be shown mathematically that with, for example, 200 offers and zero costs it is optimal to let 74 offers pass and then pick the first offer that is higher than any of the offers encountered before. The authors found that in about one-third of the cases the subjects made decision 'errors', i.e. deviations from the norm. But in roughly two-third of the cases the subjects' behavior was in agreement with the normative model.

Over time it became clearer to Tversky that normative decision theory was often too difficult to apply to actual human decision behavior. Although the normative theory was not always violated and could often be saved as descriptive theory by means of *ad hoc* assumptions, the pressing conclusion was that in too many cases people's actual decisions systematically deviate from the optimal decision as determined by measurement and decision theory. Decision theory could only be proven incorrect by *a priori* introspective reasoning (as set out in Chapter two) but persistent deviating decision behavior by people in experiments could nevertheless be an indication that something was wrong with the theory. If one observes the reasoning behavior of subjects which deviates from what is considered to be logically correct, either these subjects do not reason logically, or the assumed theory of logical reasoning is incorrect.

### *2.4 Situating Tversky's experimental method*

Tversky's increasing conviction that often the normative theory could not be applied to actual human decision behavior was the result of the experiments he conducted in the 1960s. These experiments were done with a small number of subjects; seven or eight was a normal group size. Although this was not always explicitly indicated, it is furthermore clear that in the majority of cases the experimenter and subject were previously acquainted since, for example, the subjects had participated in a university course the experimenter taught as their professor.

The experiments were done in a traditional experimental psychological setting. The subjects were often assigned numbers, for instance 1-7, and referred to individually. When discussing the empirical results, the subjects were sometimes analyzed individually, which was typically exemplary for a perceived more general

behavioral pattern. Thus, for instance, in an experiment published in 1969, the utility curve of subject 3 was discussed because of its peculiar shape, and in similar experiments conducted in 1966, statements of a post-experiment interview with subjects were compared with their performance during the experiment [Tversky and Edwards (1966), Tversky (1969)]. In this regard, Tversky's experimental work is an example of research from before the "inference revolution" of the mid-twentieth century [Gigerenzer and Murray (1987), p.182, see also Danziger (1990)]. That is to say, the analyses did not calculate an average response over experimental subjects, but instead tried to find an explanation that would cover the observed behavior of the individual experimental subjects.

The individual trials of the experiments were relatively long. For example, in one experiment [Rapoport and Tversky (1970)] subjects were asked to judge which of two light bulbs was brighter one thousand times in a row. The experiment consisted of a sequence of two or three trials of about one hour each. As a result, the experiment took quite some time. In this most time-intensive experiment done by Tversky, "[t]he subjects met five times a week for seven weeks. Each experimental session lasted about two hours" [Rapoport and Tversky (1970), p.108].

### *2.5 The road not taken: Elimination by Aspects*

One conclusion Tversky was aiming to develop from 1966-1967 onwards was that people do not behave according to the normative theory of decision making, but nevertheless act rationally [Krantz – interview (2008)]. In other words, Tversky tried to develop a new normative theory. "Elimination by Aspects: A Theory of Choice" appeared as a monograph in 1971 and as an article in *Psychological Review* in 1972. "Elimination by Aspects: A Theory of Choice" began by introducing decision theory's two-fold problem with the independence axiom. The independence axiom was first of all problematic on empirical grounds because it was "incompatible with some observed patterns of preferences which exhibit systematic dependencies among alternatives" [Tversky (1972), p.281]. Even though he had observed this problem before, Tversky now concluded that this behavioral deviation could not be solved within existing decision theory:

> data show that the principle of independence from irrelevant alternatives is violated in a manner that cannot be readily accounted for by grouping choice alternatives. More specifically, it appears that the addition of an alternative to an offered set "hurts" alternatives that are similar to the added alternatives more than those that are dissimilar to it" [Tversky (1972), p.283].

Because of the impossibility of solving the behavioral deviations within the existing theory, decision theorists and behavioral decision researchers required "a more drastic revision of the principles underlying [the] models of choice" [Tversky (1972), p.283].

In the theory as constructed by Savage (1954) only deductive, introspective reasoning could show the normative theory to be wrong. Empirical results could not show these rules of logical, rational reasoning to be false. Tversky accepted this position but moved on to show that on the basis of deductive armchair thinking doubts could also be raised. "Suppose," Tversky argued, "you are offered a choice among the following three records: a suite by Debussy, denoted D, and two different recordings of the same Beethoven symphony, denoted $B_1$ and $B_2$." Assume furthermore "that the two Beethoven recordings are of equal quality, and that you are undecided between adding a Debussy or a Beethoven to your collection. Hence, $P(B_1;B_2) = P(D;B_1) = P(D;B_2) = ½$." It then "follows readily that $P(D;B_1;B_2) = 1/3$." However, this conclusion "is unacceptable on *intuitive grounds* because the basic conflict between Debussy and Beethoven is not likely to be affected by the addition of another Beethoven recording" [Tversky (1972), p.283, emphasis added]. The empirical evidence had made the normative theory less useful for practical purposes, but it was this last introspective argument that dealt the final blow.

Thus, a new normative decision theory was required, Tversky argued, and this should preferably be a theory that could serve both the normative and the descriptive domain. Tversky then proposed such a theory, labeled elimination-by-aspects (EBA). Elimination-by-aspects was as simple and elegant as it was convincing. Rational human decision making, Tversky argued, occurs not through a process of expected utility maximization, but through a sequential process of eliminating the alternative with the lowest expected value. Here I quote Tversky at length.

> The present development describes choice as a covert sequential elimination process. Suppose that each alternative consists of a set of aspects of characteristics, and that at every stage of the process, an aspect is selected (from those included in the available alternatives) with probability that is proportional to its weight. The selection of an aspect eliminates all the alternatives that do not include the selected aspects, and the process continues until a single alternative remains. If a selected aspect is included in all the available alternatives, no alternative is eliminated and a new aspect is selected. Consequently, aspects that are common to all the alternatives under consideration do not affect choice probabilities. Since the present theory describes choice as an elimination process governed by successive selection of aspects, it is called the elimination-by-aspects (EBA) model.
> [Tversky (1972), p.285]

Tversky's EBA model is an example of attempts made by (behavioral) decision theorists in the late 1960s and early 1970s to move away from the traditional normative decision theory in order to propose alternatives. It shows that Tversky, over the course of the first eight years of his professional career, became increasingly dissatisfied with the normative theory as set out by Savage. Moreover, it shows that by the late 1960s he was actively searching for a solution in view of the difficulties, and that he sought to construct a new normative theory.

### *2.6 Caught between* a priori *axioms and behavioral deviations*

Tversky was professionally trained at the University of Michigan during the 1960s in two related traditions: the mathematical psychology of Louis Leon Thurstone, Stanley Stevens, and Clyde Coombs, and the decision theory/behavioral decision research of Leonard Savage and Ward Edwards. In the work of Tversky these two branches of psychology came together, extending and influencing each other. The problem that arose was that the experiments produced systematic behavioral deviations that potentially disproved the very foundations of measurement theory and decision theory. Because of the far-reaching consequences of accepting that the axioms were falsified, Tversky was reluctant to accept this conclusion. At the same time, however, he took the behavioral deviations seriously and was unwilling to accept as an explanation for the behavioral deviations *ad hoc* explanations that problematized the

experimental procedure, such as the solution of changing tastes between experimental sessions. By the late 1960s and early 1970s, Tversky was actively looking for a way to solve the behavioral deviations problem. His eliminations-by-aspects theory was an attempt to take the behavioral deviations seriously, and to reconstruct decision theory and measurement theory based on a new foundation.

The methodological tension Tversky was struggling with involved how a theory that is *a priori* true can be combined with experimental results that point in many directions, but only occasionally in the direction of the *a priori* theory. Tversky could not proceed as an experimentalist who seeks to test whether a theory is right or wrong. As the axioms were *a priori* truths, they could only be proven wrong on the basis of *a priori* reasoning. Measurement theory and normative decision theory were simply not devised and employed as theories that could be proven wrong experimentally. At the same time, however, Tversky wanted to do justice to the experimental results he had obtained. Sticking to the axioms of measurement theory and decision theory would have implied that whenever a systematic behavioral deviation was observed in the experiments, there was something wrong with the experiment. This would mean that the majority of Tversky's experiments were invalid. Tversky was thus effectively caught between the *a priori* truth of the axioms of measurement theory and decision theory, and the behavioral deviations that surfaced in his experiments. He had to decide between either taking his experiments seriously, or accepting the axioms of measurement theory and decision theory. His elimination-by-aspects theory proves that by the early 1970s, Tversky had chosen the first option. He accepted his experimental results as valid and thus had to construct a whole new basis for measurement theory and decision theory. But he did so by disproving Savage's decision theory on intuitive grounds. That is, by accepting that ultimately only an *a priori* 'test' of the axioms could prove them wrong.

For reasons that will become clear in Chapter four, the elimination-by-aspects theory turned out to be a road not taken. In the beginning of the 1970s, Kahneman offered Tversky a solution that both solved the problem of experimental behavioral deviations, and at the same time left intact the fundamentals of measurement theory and decision theory. The Savage-Edwards approach to decision theory and behavioral decision research continued to be problematic for Tversky, but the solution he would come to favor lay in another direction. Tversky would find this solution in his joint work with Kahneman.

## 3. Kahneman's cognitive mistakes

### *3.1 From correlational to experimental psychology*

Kahneman obtained a B.A. from Hebrew University in 1956 while working as a psychologist in the Israeli army. In 1958 he moved to San Francisco and obtained a PhD from the University of California at Berkeley in 1961 under the supervision of Susan Ervin (1927- ). To understand how Kahneman solved Tversky's problems in mathematical psychology, decision theory and behavioral decision theory in the 1970s, we need to examine the research Kahneman conducted before his collaboration with Tversky. Between 1961 and 1971, Kahneman's research was about semantic differentials, optometry, vision research, and related themes. Kahneman's research in this period was unrelated to mathematical psychology, unrelated to decision theory, unrelated to behavioral decision research, and despite some retrospective hints of Kahneman to the contrary, entirely unrelated to economics.

Based on Kahneman's recollections in his autobiography and the one publication that emerged from this, his early work for the Israeli army in the early 1950s and at the Hebrew University is best characterized as correlational psychology [Danziger (1990,1997), Gigerenzer (1987a,b)].[21] Correlational psychology builds theories on the basis of correlations in statistical data, for example between IQ and the degree of education. Using methods developed by the British army in World War Two, the aim of Kahneman's early research was to develop reliable predictions about the future performance of people on the basis of character traits, be it in the army or in different kinds of jobs. For instance, to find out at an early stage which new recruits in the army would eventually be successful leaders on the future battlefield, different tests were designed to evaluate the differences between recruits with respect to a few behavioral and personal characteristics that were thought to relate to leadership capacities.

It is not difficult to see that in this kind of research the ability of the researcher to predict the future performance of the subjects investigated is an important, and perhaps the only way to measure success. A classification of new recruits in the army along different dimensions might be an interesting exercise, but if it does not predict

[21] Kahneman and Ghiselli (1962), Kahneman (2002). After he completed his PhD at Berkeley in 1961 Kahneman returned to the psychology department at Hebrew University where he would remain until 1978. In the meantime, however, he was a visiting scholar at the University of Michigan in 1965-1966, a visiting scientist and lecturer in psychology at Harvard University in 1966-1967, a visiting scientist during the summer terms of 1968 and 1969 at the Applied Psychology Research Unit of Cambridge, UK, and a lecturer in the graduate program of the University of Michigan in 1968-1969.

better than chance then it is of no use. In his autobiography [Kahneman (2002)] Kahneman recalls how frustrating it was when time and again he was confronted with the fact that his predictions were anything but reliable. Extensive questionnaires and tests were set up, but in the end it turned out that the intuitive guesses of the staff members who conducted the tests and collected the questionnaires proved better than the scientific predictions.

Dissatisfied with the results of this research and eager to develop his research skills, Kahneman switched to experimental psychology of vision, resulting in some twenty-five articles over a period of ten years, including two publications in *Science*, and a whole range more in prominent experimental psychology journals such as the *Journal of Experimental Psychology*. There is not one particular theme or article that stands out during the decade from 1961 to 1971. The psychological view held by Kahneman emerges when the different themes and articles are considered next to each other. In the following two sections I will first provide an overview of Kahneman's work during this period and then come to a general assessment of Kahneman's work in this period.

### *3.2 Kahneman's research 1961-1971*

Kahneman's career began with the theoretical work he did concerning the models used in experimental studies of semantic differentials. Semantic Differential (SD) research investigates people's attitude towards words [Heise (1970), p.235], or, put differently, measures the meaning of abstract objects to the individual [Kiddler (1981)]. A distinction is made between the denotative and the connotative meaning of a word, or concept, or object. Thus, it is assumed that apart from the dictionary, or the denotative meaning, words are also assigned connotative meanings by the individual. The words *massacre* and *rape*, for instance, are attributed different connotations than *flower* or *sunny day* in terms of good versus bad. SDs are measured on a bipolar scale, for which in principle all opposites can be used, such as good-bad, soft-hard, fast-slow, clean-dirty, valuable-worthless, and so forth.

Articles on SD make up a small, but important part of the publications of Kahneman's early work. Apart from his dissertation, it was the subject of one published article.[22] It also provides a good illustration of Kahneman's take on

[22] Kahneman's dissertation consisted of a paper he wrote in eight days [Kahneman (2002), p.6].

experimental psychology. Within the field of SD research, Kahneman's focus was on the theory behind the models that are used to infer conclusions about the connotative meanings. In Kahneman (1963) he showed that models that are used to measure SDs are mathematically not sufficiently sophisticated and may give rise to wrong interpretations of what is observed. Kahneman considered the following model for the rating $s_{ijk}$ of the concept $j$ by individual $i$ on scale $k$

$$s_{ijk} = T_{jk} + C_{ik} + d_{ijk} \qquad [1]$$

in which $T_{jk}$ is the "true score" of concept $j$ on scale $k$, computed as the average score of a number of judges. $C_{ik}$ is the "constant deviation" of subject $i$ on scale $k$, computed as the average deviation of subject $i$ from the true score over a large number of concepts. $d_{ijk}$ is the "specific deviation" (or "error of judgment") on a particular rating. Kahneman argued that the practice of contemporary SD research wrongly assumed that the specific deviations of ratings were uncorrelated. For example, deviating from the true score could very well be correlated during the course of one experiment. Improvement should be sought in the direction of more "precise algebraic" models. Kahneman's research focused not so much on the theory of SDs as such, but on the improvement of the analysis of variance in the statistical models it employed. Specifically, he focused on the notion of the "error of judgment" in SD research. The object of investigation was to understand how people deviate from what is true or correct. In effect, that meant that the analysis of the actual process of how people make judgments was black-boxed.

In 1962-1963, Kahneman set up a vision lab at the department of psychology of the Hebrew University [Kahneman (2002), p.6]. Many of the articles he published in the following years were derived from the experimental results of this lab. In this research, Kahneman investigated the relationship between the "energy" of different stimuli and visual perception capacities. "Energy" was employed as a general concept to define the strength of a stimulus; the brighter, the more illuminated, the more contrasted, the longer and so forth the stimulus was, the more energy it had. Visual perception was measured in terms of the reaction times of the subjects. In the typical experiment, the subject had to decide as quickly as possible whether the opening of a

so-called Landolt C was directed up-, down-, left-, or rightwards.[23] The conditions in terms of brightness, contrast, and so on in this setting could be varied in numerous ways. The visual task could also be combined with other cognitive tasks. Kahneman's text book on the psychology of vision and attention, *Attention and Effort* (1973), is still used in the early twenty-first century as standard reference on the subject [Dawes – Interview (2008)].

Examples of this research include Kahneman's (1964), "Control of Spurious Association and the Reliability of the Controlled Variable" and Kahneman's (1966b), "Time-Intensity Reciprocity in Acuity as a Function of Luminance and Figure-Ground Contrast."[24] In Kahneman and Norman (1964), the relation between the minimal amount of time subjects need to identify a visual stimulus (labeled the "critical duration" $t_c$) and the energy in terms of brightness and duration of the stimulus is investigated. It was shown that the stimuli of equal energy do not necessarily produce the same critical duration and that a given visual stimulus does not trigger one but multiple sensory processes. The second conclusion particularly opposed the general view held in the psychophysical community that one stimulus triggers only one sensory process. In Kahneman (1966b) the Bunsen-Roscoe law of the time-intensity of reciprocity is (partly) falsified. This law forms a central concept in the psychophysics of vision, stating that up to the critical duration $t_c$, duration and intensity of the stimulus are interchangeable, in which duration is measured in seconds and intensity in lux. Kahneman showed that time-intensity reciprocity fails to hold when a Landolt C stimulus at 40mL*msec is preceded by a 2 second flash of 1mL.

In the psychophysical paradigm, visual perception is seen as one of many cognitive tasks. Other cognitive tasks include conversation, or more generally, speech, learning, and calculation. How different cognitive tasks influence one another was investigated in Kahneman and Beatty (1966, 1967), Kahneman et al. (1967,1968), and Kahneman and Peavler (1969). The explicit emphasis in these articles was on how the combination of different cognitive tasks could lead to "errors of judgment." In Kahneman et al. (1967) for instance it was shown that the capacity to visually

[23] The Landolt C is one of the standard symbols used in American psychophysics of vision and optometry. It consists of a C in which the opening can be varied, and which is either surrounded by bars the width of which equals the C's opening or not surrounded.
[24] Other examples include Flom et al. (1963), Kahneman (1965a,b, 1966a, 1967), Kahneman and Norman (1964), and Kahneman et al. (1967).

perceive substantially decreases when subjects were engaged in other mental tasks such as speech or calculation. The "error of judgment" in these cases is very real, as it explains for instance why car drivers may miss a stop sign when engaged in conversation. It again illustrates Kahneman's focus on the psychology of mistakes.

Another way in which the psychophysics of vision and cognitive tasks are connected is through measuring a mental task. For example, a clear correlation can be found between the difficulty of the mental task and the diameter of the pupil. The pupil dilates when the task begins and constricts when the answer or report is given. This relation is investigated in Kahneman and Beatty (1966), Kahneman et al. (1968), Kahneman et al. (1969), and Kahneman and Wright (1971). In Kahneman and Wright (1971), the authors investigated the correlation between pupil size and short-term memory. It was shown that pupillary size provides a good measurement for mental activity. When involved in a mental task, the subjects' pupil clearly dilates. A number of characteristics of short term memory were reported. Mental activity in case of short term memory seemed for example negatively correlated with the time subjects have to store a stimulus in memory. If a stimulus has to be stored in memory for seven seconds, instead of three, mental activity on average was lower. However, there are more factors influencing the level of activity. It was also shown that mental activity and the difficulty of a memory task are positively correlated. Kahneman et al. (1969) compared the pupil diameter as a measurement for mental activity with measurements of the heart rate and of skin resistance. The results indicated that these three measurements are not very well correlated. Pupil diameter remained therefore the preferred method. Nevertheless, some reservation was expressed, as the relation between mental activity and each of the three methods is not precisely known.

### *3.3 Kahneman's cognitive errors*

In Kahneman's vision research an emphasis was placed on the question under which circumstances the human mind makes cognitive errors. Kahneman showed that there is a trade-off between different cognitive tasks in perception capacities, and that as a result people may sometimes "fail" to perceive the stimulus and make an error in judgment. Furthermore, the research done by Kahneman in the period between 1961 and 1971 was in line with the interwar drive to eliminate all introspection from psychology [Danziger (1997)]. In Kahneman's experiments self-reports were not necessary to establish how the cognitive system operates. The behavior of the

cognitive system could be inferred from observed behavior and physical responses which cannot be controlled, such as pupil dilations and restriction. The human mind was considered to be a black box whose functioning could be inferred from the uncontrollable and unconscious responses made by the individual subjects.

Both elements are important in gaining an understanding of Kahneman's psychology and his subsequent influence on Tversky. The recurring theme of the cognitive errors shows, that in Kahneman's view, psychology was about discovering how people deviate from a norm behavior. This aspect of experimental psychology dates back to the beginning of experimental psychology in nineteenth-century Germany. But in nineteenth-century German and interwar American experimental psychology, this framework was adopted for the purpose of discovering what the *true* value was. The experimental psychologists wanted to know the true value of, for instance, the smallest amount of difference in weight people could perceive, and for this purpose devised a framework, which in spite of all the individual errors, could establish the true value. Thurstone, for instance, wanted to measure *the* attitude towards religion of a group of people, and for this purpose he constructed a method that would elicit the attitude from a series of observations in which each individually deviated from the true value. Experimental psychology was explicitly modeled after experimental practice in physics, where the physicist tries to establish the true value of boiling water by conducting a series of measurements in which each measurement individually deviates from the true value and from each other.

Kahneman employed the experimental psychological framework, but applied it differently than the nineteenth-century German psychologists. In Kahneman's work the true value was known. The true value was an accurate prediction of a recruit's future leadership capacities, or the true value was not running through a traffic light when driving a car. The question Kahneman then raised was how, when, and why the cognitive machinery fails to act according to the true value. Kahneman used an experimental psychological framework, but applied it with the opposite purpose. He did not want to find out what the *true* value was, but how people *deviate* from the true value. In Kahneman's research, the true value was always clear and determined by the experimenter. Kahneman knew how the cognitive machinery ideally responds, and investigated whether it actually does do so. In Kahneman's understanding, the scientist thus completely determined in each experimental situation what the good, optimal, or rational behavior should be. This was in line with the scientific desire to

eliminate all introspection because it assumed that the experimental subject cannot judge whether he or she is giving the correct response or not. In Kahneman's experiments the experimenter determined how the subject should behave and determined how it did behave. All authority for judging behavior was placed in the hands of the scientist.

Because Kahneman has never provided an extensive theoretical exposition of the assumption that human beings often make cognitive errors, one could easily dismiss it as merely a nice way of illustrating theories which are perhaps not too exciting, but that would be a mistake. The key to understanding Kahneman's psychology lies in his conviction that human beings often make cognitive errors. Since Kahneman is a psychologist who sets out his theories through case-based reasoning, it is also by means of these cases that we can best illustrate and understand Kahneman's firm belief in cognitive errors. In the example of the army psychologist at the beginning of this chapter, Kahneman and his colleagues really believed that through their extensive studies they could accurately predict, or at least predict better than by mere chance, the future performance of different candidates for a job. The fact that they could not was for the young Kahneman a true cognitive illusion that he needed to correct for himself [Kahneman (2002)].

Another illustrative example recalled by Kahneman in his autobiography was the moment a flight instructor disagreed with the psychologists' theory that praise is more effective in developing skills than punishment. The flight instructor reasoned that although he praised the good performance of his recruits, the next time the performance would almost always be worse. Similarly, he would always punish recruits who had done a poor job, and this would almost always improve performance the next time. To Kahneman this was a clear cognitive illusion. A good performance is statistically more likely to be followed by a worse performance than by an equally good or even better performance, and vice versa. Thus, the flight instructor was suffering from a cognitive illusion. The truck or car driver described above who is engaged in a conversation and thus does not see a traffic light that he or she would otherwise not miss, really does makes an error. His or her cognitive apparatus is tuned to noticing traffic lights, but it fails to do so.

To Kahneman it was and is a given fact of life that human beings often make cognitive errors. However, science can help in two ways. First, scientists can set out what the correct way of behaving should be for each situation. For the truck driver, it

is obvious what the correct behavior is, but for the flight instructor it may not be intuitively clear what the correct way of reasoning is. Scientists can therefore help to establish the correct way of reasoning. Second, scientists, and in particular psychologists, can help by investigating when, how and in what way human beings make cognitive errors and thus provide a basis for designing tools or education to help human beings correct these cognitive errors.

**4. Conclusion**

Before Kahneman and Tversky started collaborating in 1969 and published their first paper in 1971, the research that each conducted was not directly related. Although both can, partly in the case of Tversky, be placed in the realm of experimental psychology, the approach each took and the psychological phenomena they investigated differed. Tversky was raised in the fields of mathematical psychology, decision theory, and behavioral decision research at the University of Michigan. He was principally interested in human decision making under risk. Kahneman, on the other hand, conducted research on semantic differentials, vision and the interaction of cognitive tasks.

The point of contact between their two research programs was that of experimentally observed behavior that deviates from what the scientist expects. Tversky was struggling to incorporate behavioral deviations in the representational theory of measurement and decision theory, whereas Kahneman was investigating how, when, and why human beings deviate from the response or behavior they should display. This concern with deviating behavior would become the basis for their collaborative research of the 1970s.