

Psychologie, experimentelle Ökonomie und die Frage, was gutes Experimentieren ist

Ralph Hertwig

Max-Planck-Institut für Bildungsforschung, Berlin

Zusammenfassung. Die Beziehung zwischen Psychologie und Ökonomie zu charakterisieren, hieße von „Beziehungslosigkeit“ zu sprechen. Ausgenommen von seinem Urteil hat Bruno Frey (1990, S. 67) lediglich den Bereich der Entscheidungsforschung, genauer das Forschungsprogramm von Amos Tversky und Daniel Kahneman. Tenor dieses Programms ist es, daß Menschen systematisch diverse Normen verletzen, die traditionell als konstitutiv für Rationalität erachtet werden. Diese Normverstöße, sogenannte kognitive Illusionen, widersprechen dem ökonomischen Modell menschlichen Verhaltens, denn dieses setzt rationales Verhalten im Sinne dieser Normen voraus. Experimentelle Ökonomen haben diese Herausforderung durch die Psychologie mit einer Methodenkritik erwidert, die die Validität dieser psychologischen Befunde in Frage stellt: Im Unterschied zu vielen Psychologen verwenden experimentelle Ökonomen viel Mühe darauf, die Versuchsperson zu motivieren, ihr die Möglichkeit zu geben zu lernen und sie nicht zu täuschen. Es gibt gute Gründe, warum Psychologen diese methodischen Präferenzen in Betracht ziehen sollten.

Schlüsselwörter: Experimentelle Ökonomie, kognitive Täuschungen, experimentelle Methoden, leistungskontingente Bezahlung, Täuschung

Psychology, experimental economics and the question of what good experimentation is

Abstract. Research in psychology and economics rarely interacts. A notable exception is Amos Tversky and Daniel Kahneman's research on decision making. The gist of their program is that humans systematically violate norms that have been equated with rationality. These so-called cognitive illusions cast doubt on the rationality assumption of economic models of human behavior. Economists have questioned the validity of these results by criticizing the experimental methods used in psychology. Unlike many psychologists, experimental economists seriously attempt to motivate their participants, provide them with the opportunity to learn, and avoid deceiving them. There are good reasons why psychologists ought to consider those methodological preferences.

Key words: experimental economics, cognitive illusions, experimental methods, performance-contingent payment, deception

Einführung

Ökonomen besitzen einen gesellschaftspolitischen Gestaltungsanspruch, der sie von vielen anderen Sozial- und Geisteswissenschaftlern unterscheidet. Paul

Ich möchte mich für hilfreiche Diskussionen des Manuskripts bei Gerd Gigerenzer, Ulrich Hoffrage, Andreas Ortmann und Peter Sedlmeier sowie für viele Verbesserungsvorschläge bei Jürgen Baumgarten, Ursula Dohme, Karl Christoph Klauer, Erika Nüssle, Erna Schiwietz, Jutta Wolf und einem anonymen Reviewer herzlich bedanken.

Samuelson, Nobelpreisträger in Ökonomie, hat diesen Gestaltungsanspruch pointiert in Worte gefaßt: „I don't care who writes a nation's laws ... if I can write its economics textbooks“ (zitiert in *The Economist*, 1997, S. 13). Wer sich derart exponiert, muß mit Widerrede rechnen. Trotzdem kommt sie überraschend, wenn ihr Forum ausgerechnet das *Wall Street Journal* ist. Mit deutlichem Sarkasmus spießte Loewenstein (1996) dort die theoretischen Grundüberzeugungen vieler Ökonomen auf:

To mainstream economists, markets are the great citadel of rational man. When buying a car, you and I may indulge a fantasy, or a neurosis. But as participants in impersonal markets [d. h. in Finanzmärkten; R. H.], we are cool, calcu-

lating and predictable According to the efficient-market theory, which is the state religion of economics departments and business schools, prices are so *uniformly* rational that no one can consistently profit from exploiting mis-priced stocks. (S. 1)

Loewenstein schrieb diese Sätze aus Anlaß des Todes von Amos Tversky, des wohl einflußreichsten psychologischen Entscheidungsforschers der letzten 25 Jahre und scharfen Kritikers der ökonomischen Entscheidungstheorie (vgl. Rabin, in Druck). Die *conditio sine qua non* der ökonomischen Entscheidungstheorie ist, daß Menschen sich rational – im Sinne der Maximierung des erwarteten Nutzens (von Neumann & Morgenstern, 1944) – verhalten. Die Arbeiten von Tversky und seines Kollegen Daniel Kahneman scheinen hingegen zu demonstrieren, daß – in den Worten der *New York Times* – „people do not always behave rationally when they make decisions ... and that there are many more quirks in the human reasoning process than many earlier economic and psychological theories had contended“ (Freeman, 1996, B16).

Wie haben Ökonomen auf diese Herausforderung an die Theoriebildung in der Ökonomie reagiert? In *Teil I* dieses Artikels werden zwei Reaktionen berichtet. Besonders relevant ist sicherlich diejenige, die die deutlich unterschiedlichen experimentellen Praktiken in der Psychologie und in der Ökonomie konstatiert und vermutet, daß die von Tversky und Kahneman dokumentierten Normverstöße an deren experimentellen Praktiken gebunden seien. In *Teil II* des Artikels werde ich am Beispiel von drei methodischen Präferenzen die experimentellen Methoden der Ökonomie illustrieren und argumentieren, daß es gute Gründe gibt, warum diese auch in der Psychologie Anwendung finden sollten.

Teil I: Tversky und Kahnemans Herausforderung an die Theoriebildung in der Ökonomie

Seit den frühen 1970er Jahren akkumuliert das einflußreichste Forschungsprogramm der psychologischen Entscheidungsforschung, das Heuristics-und-Biases-Programm, systematische Fehler in menschlichen Urteilen unter Unsicherheit (z. B. Kahneman & Tversky, 1979; Tversky & Kahneman, 1981; Kahneman, Slovic & Tversky, 1982; Tversky & Kahneman, 1987; Kahneman & Tversky, 1996). Die Strategie des Programms ist einfach: In Gestalt von Gesetzen der Wahrscheinlichkeitstheorie und der Erwartungsnutzentheorie (sowie Axiomen, auf denen jene aufgebaut) werden normative Kriterien als Nullhypothese über Verhalten (d. h. Präferenzen, Urteile, Entschei-

dungen) postuliert.¹ Wenn beobachtetes Verhalten normative Kriterien systematisch verletzt, dann wird die Nullhypothese verworfen. Anlässe dafür gab und gibt es viele. Allein der Katalog von Verstößen gegen probabilistische Gesetzmäßigkeiten umfaßt unter anderem *base rate neglect*, *law of small numbers*, *misperception of random numbers* und *conjunction fallacy*. Bestandsaufnahmen dieser Forschung aus der jüngsten Zeit findet man zum Beispiel bei Mellers, Schwartz und Cooke (1998).

Einige Ökonomen haben früh auf die Existenz dieser Befunde reagiert. Bereits 1978 veröffentlichte David Grether im *American Economic Review* eine Zusammenfassung von „Recent psychological studies of behavior under uncertainty“, und ein Jahr später machten er und Charles Plott (1979) ihre Kollegen auf das *Preference-reversal*-Phänomen aufmerksam:

A body of data and theory has been developing within psychology which should be of interest to economists. Taken at face value the data are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency ... suggests that no optimization principles of any sort lie behind even the simplest of human choices and that the uniformities in human choice behavior which lie behind market behavior may result from principles which are of a completely different sort from those generally accepted. (S. 623)

Es war gleichfalls Grether (1980), der die Studie von Kahneman und Tversky (1973) zur Mißachtung von Basisraten unter den methodischen Präferenzen der experimentellen Ökonomie zu replizieren suchte.²

¹ Neben diesen Kriterien fanden zum Beispiel auch objektive Häufigkeiten als Normen Verwendung (Tversky & Kahneman, 1973).

² In Kahneman und Tverskys (1973, S. 241) Experiment zur Mißachtung von Basisraten lasen die Versuchspersonen die folgende Instruktion: „A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale of 0 to 100. The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.“ Zwei Gruppen von Versuchspersonen erhielten diese Instruktion. Allerdings erfuhr eine Gruppe, daß es sich um 30 Rechtsanwälte und 70 Ingenieure handle, und für die andere Gruppe wurden die Basisraten vertauscht. Trotz der vertauschten Basisraten waren die gesuchten a-posteriori-Wahrscheinlichkeitsurteile in beiden Gruppen beinahe identisch. Der Befund wurde von Kahneman und Tverskys (1973) als Beleg dafür gewertet, daß ihre Versuchspersonen die Basisraten vernachlässigt haben.

Anhand dieses Phänomens läßt sich veranschaulichen, warum die Befunde des Heuristics-und-Biases-Programms für die Theoriebildung in der Ökonomie brisant sind. Eine klassische Annahme ökonomischer Modelle ist, daß Marktteilnehmer Wahrscheinlichkeitsurteile mit Hilfe des Bayes-Theorems revidieren. Die Annahme ist deshalb äußerst nützlich, weil – sollte der Markt so funktionieren, als ob zum Beispiel Händler „Bayesianer“ sind – ökonomische Theorien dann ein bestimmtes Muster von Preisen und Allokationen der Ressourcen vorhersagen können.

Ob Händler in einem Markt tatsächlich „Bayesianer“ sind, kann mit Hilfe eines simulierten experimentellen Marktes untersucht werden, der wie folgt organisiert sein könnte (siehe experimentelle Realisationen zum Beispiel bei Duh & Sunder, 1986; Camerer, 1987, 1990): Händler handeln Aktien, für die sie am Ende des Marktgeschehens Dividenden erhalten. Die Dividende einer Aktie und damit auch ihr Wert hängen davon ab, ob (1) am Ende des Marktgeschehens das Eintreten von Ereignis X oder Y bekanntgegeben wird, und ob (2) die Gewinnausschüttung einer Aktie mit dem Eintreten des Ereignisses X oder Y gekoppelt ist. Um ihren Gewinn zu maximieren, müssen die Händler die Auftretenswahrscheinlichkeit der Ereignisse X und Y abschätzen. Beurteilt ein Händler das Eintreten von X als wahrscheinlicher, dann sollte er X -Aktien kaufen und Y -Aktien verkaufen.

Die Ereignisse X und Y können durch zwei Urnen repräsentiert werden: Urne X enthält 1 rote und 2 schwarze Kugeln und Urne Y enthält 2 rote und 1 schwarze Kugel. Per Zufall entscheidet der Experimentator zu Beginn, ob am Ende des Marktgeschehens Ereignis X oder Y bekanntgegeben wird. Der Händler wiederum erfährt zu Beginn nicht, welches Ereignis eintreten wird, aber er erfährt die Basisrate der X - und Y -Ereignisse (z.B. .6 und .4) und eine „Likelihood“-Information: Angenommen, Ereignis X (d.h. Urne X) wurde ausgewählt, dann werden aus Urne X 3 Kugeln (mit Zurücklegen) gezogen, und die Komposition dieser Stichprobe, aber nicht deren Herkunft, wird den Händlern mitgeteilt. Eine Stichprobe von 1 roten und 2 schwarzen Kugeln deutete auf Ereignis X , und ein „Bayesianischer“ Händler könnte die *a-posteriori*-Wahrscheinlichkeit für X (gegeben 1 rote Kugel) wie folgt berechnen:

$$p(X | 1 \text{ rot}) = p(1 \text{ rot} | X) p(X) / (p(1 \text{ rot} | X) p(X) + p(1 \text{ rot} | Y) p(Y)).$$

Wenn man den Befunden von Kahneman und Tversky (1973) folgt, dann sind Bayesianische Urteile kaum zu erwarten. Eher würde man erwarten, daß Händler in einem solchen Markt die Repräsentativitäts-Heuristik anwenden. Danach würde das Aus-

maß der Korrespondenz zwischen dem Verhältnis der schwarzen und roten Kugeln in der Stichprobe und dem in der Urne als Index für deren Auftretenswahrscheinlichkeit verwendet („Representativeness is an assessment of the degree of correspondence between a sample and a population“, Tversky & Kahneman, 1983, S. 295). Wäre dies der Fall, dann würden sich die Preise und Allokationen der Ressourcen von denen unterscheiden, die man für einen „Bayesianischen“ Markt vorhersagt. Brisant wäre dies deshalb, weil der Markt sich dann entgegen der Annahme der ökonomischen Modelle nicht optimal verhielte.

Eine Revision von Wahrscheinlichkeiten, die nicht dem Bayes-Theorem folgt, ist nur ein Beispiel für einen Verstoß gegen traditionelle Annahmen ökonomischer Theoriebildung. Einen Katalog dieser Verstöße und der ökonomischen Überzeugungen, mit denen sie in Konflikt stehen, kann man unter anderem bei Thaler (1987a), Camerer (1995) und Rabin (in Druck) finden. Alle drei Arbeiten sind auch deshalb von Bedeutung, weil sie stellvertretend für den Versuch der „school of behavioral economics“ stehen, Theoriebildung in der Ökonomie auf die Basis einer systematischen Theorie von Urteilsfehlern zu stellen. Wie reagieren Ökonomen auf die Existenz „kognitiver Illusionen“ (dieser Ausdruck wird hier in einem weiten Sinn gebraucht und bezieht sich auf Verstöße gegen die Wahrscheinlichkeits-, die Erwartungsnutzentheorie etc.)?

Wie wurde die Rationalitätsannahme verteidigt?

Es gibt eine Vielzahl von Argumenten, mit denen Ökonomen ihre Theorien gegen den Vorwurf verteidigen, daß Individuen sich nicht rational verhalten. Eine Bestandsaufnahme der gängigen Argumente findet man zum Beispiel bei Grether (1978), Camerer (1987), Frey (1990) und Roth (1996). Im folgenden werden zwei Argumente erörtert: das Markt-Argument und das Artefakt-Argument.

Markt-Argument

Vom vermutlich wichtigsten theoretischen Argument gibt es verschiedene Varianten, denen die Annahme gemeinsam ist, daß aggregiertes Verhalten, das heißt das Verhalten des Marktes, nicht mit dem Verhalten von Individuen gleichzusetzen ist. Weil diese Gleichsetzung nicht gilt, so das Argument, können Theorien, die auf falschen Annahmen aufbauen (z. B. der Annahme, daß sich Individuen rational verhalten),

trotzdem den Markt gut vorhersagen (Friedman, 1953).

Marktverhalten kann sich aus diversen Gründen vom Verhalten individueller Agenten unterscheiden (siehe Camerer, 1990). Zum Beispiel könnte es zutreffen, daß Fehler in Entscheidungen individueller Akteure zufällig verteilt sind und sich daher im Aggregat aufheben. Oder es könnte sein, daß jene Akteure, die systematische Fehler machen, in Wettbewerbsmärkten verdrängt werden (d.h., es genügen wenige rationale Agenten, um die irrationalen vom Markt zu verdrängen). Es könnte auch sein, daß die weniger rationalen Akteure mit der Zeit von den rationalen Akteuren lernen (implizit oder explizit durch z.B. Consulting). Schließlich wäre es auch möglich, daß Akteure in Märkten ausreichend finanziellen Anreiz und Erfahrung haben, um jene Fehler zu vermeiden, die in psychologischen Experimenten gefunden werden (hier zeichnet sich schon ein Argument ab, das uns später in Gestalt des Artefakt-Arguments wieder begegnen wird). Widerspruch gegen das Markt-Argument in theoretischer und empirischer Form kann man zum Beispiel bei Russell und Thaler (1985), Akerlof und Yellen (1985) und Camerer (1992) nachlesen.

Das Markt-Argument ist auch deshalb aufschlußreich, weil es eine unterschiedliche Einstellung zur Rolle des Experiments markiert. Seit Wilhelm Wundt 1879 das erste experimental-psychologische Labor in Leipzig gründete, war und ist das Experiment in der Psychologie die Methode der Wahl zum Erkenntnisgewinn.³ Im Gegensatz dazu hielten Ökonomen (und manche halten noch immer) Laborexperimente lange Zeit für unmöglich – unter anderem deshalb, weil ihr Erkenntnisobjekt, das Marktgeschehen, nicht das Verhalten eines Einzelnen darstellt. Vielmehr resultiere dieses Geschehen aus dem sich gegenseitig bedingenden Verhalten von vielen und ließe sich daher experimentell nicht untersuchen.⁴

³ Bereits 1863 beschreibt Wilhelm Wundt in seinen *Vorlesungen über die Menschen- und Thierseele* die zentrale Rolle des Experiments für die moderne Psychologie. „... daß Experiment in der Psychologie das Haupthilfsmittel ist, welches uns von den Thatsachen des Bewußtseins auf jene Vorgänge hinleitet, die im bunten Hintergrund der Seele das bewußte Leben vorbereiten So leitet uns immer und überall erst das Experiment zu den Naturgesetzen, weil wir nur im Experiment gleichzeitig die Ursachen und die Erfolge zu überschauen vermögen.“ (S. v-vi)

⁴ Joan Robinson (1977) zum Beispiel schrieb, daß „economists cannot make use of controlled experiments to settle their differences; they have to appeal to historical evidence“ (S. 1319), und Russell und Wilkinsons (1979) Diktum war, daß „since economics is a social rather than a laboratory science, however, the opportunities to test the predictions of economic theory are limited by the experi-

Artefakt-Argument

Ebenso wie das Markt-Argument konzidiert auch das Artefakt-Argument Normverletzungen des *individuellen* Akteurs. Ersteres bezweifelt deren Generalisierbarkeit auf den Markt. Letzteres bezweifelt deren Generalisierbarkeit auf jene experimentellen Bedingungen, die Ökonomen als relevant erachten (siehe z.B. Grether & Plott, 1979).

Die Generalisierbarkeit wird zum Beispiel dadurch beeinträchtigt, daß Akteure im Marktgeschehen, anders als Psychologiestudenten im Labor, motiviert sind, fehlerhafte Entscheidungen zu vermeiden. Ergo gilt: Will man ökonomische Theorien oder die Rationalitätsannahme, auf der sie basieren, testen, dann ist der Test nur dann fair, wenn die Bedingungen, unter denen die Theorien Gültigkeit haben, auch im Labor realisiert sind. Das typische psychologische Experiment genügt vermutlich eher selten den „ökonomischen“ Testbedingungen. Grether und Plott (1979) haben die dadurch genährten Zweifel an der Generalisierbarkeit psychologischer Befunde pointiert zum Ausdruck gebracht. Sie schlugen insgesamt 13 Erklärungen für das Zustandekommen des Preference-reversal-Phänomens vor. Einige dieser Erklärungen thematisieren ausschließlich die Unterschiede des Experimentierens von Psychologen und Ökonomen. In der Erklärung „Misspecified Incentives“ zum Beispiel argumentierten die Autoren, daß

almost all economic theory is applied to situations where the agent choosing is seriously concerned or is at least choosing from among options that in some sense matter Theories about decision-making costs do suggest that unmotivated choice behavior may be very different from highly motivated choice behavior, but the differences remain essentially unexplored. Thus the results of experiments where subjects may be bored, playing games, or otherwise not motivated, present no immediate challenges to theory. (S. 624)

Grether und Plott (1979) praktizierten höfliche Kollegialität und wiesen sorgfältig darauf hin, daß „psychologists whose work we are reporting are careful scientists“ (S. 628). Allerdings, „yet a bit of suspicion always accompanies a trip across a disciplinary

ments performed by history“ (S. 5). Hier offenbart sich eine historische Parallele zu einer ähnlichen Problematik in der Psychologie. Danziger (1990) schreibt, daß Wundt, selbst in den Momenten, in denen er der experimentellen Methode am stärksten zugeneigt war, eine andere, nicht-experimentelle Psychologie, die *Völkerpsychologie*, für notwendig erachtete. Da höhere mentale Prozesse des Individuums so variabel sind wie sie sind, schlug er zu deren Studium die vergleichende Analyse von kollektiven Kulturprodukten (wie z.B. Sprache, Mythen und Gebräuche) vor: Nur im kollektiven Ausdruck können höhere mentale Prozesse die Stabilität gewinnen, die für deren Studium notwendig ist.

boundary“ (S. 628). Bei Grether und Plott geht das Mißtrauen immerhin so weit, daß sie in der „The Experimenters were Psychologists“-Erklärung die Experimentatoren selbst zur unabhängigen Variablen machten. Sie schreiben, daß

subjects nearly always speculate about the purposes of experiments and psychologists have the reputation for deceiving subjects. It is also well known that subjects' choices are often influenced by what they perceive to be the purpose of the experiment. In order to give the results additional credibility, we felt that the experimental setting should be removed from psychology. (p. 629)

Ist Grether und Plotts Mißtrauen gerechtfertigt?

Teil II: Bringt „gutes“ Experimentieren die Resultate des Heuristics-und-Biases-Programms zum Verschwinden?

Es gibt eine Reihe von Unterschieden zwischen den experimentellen Praktiken in der Psychologie und der Ökonomie. Grether und Plotts Liste ist keineswegs vollständig, aber sie enthält die Unterschiede, die in diversen Vergleichen zwischen den experimentellen Praktiken beider Disziplinen häufig genannt werden (z.B. Cox & Isaac, 1986; Lopes, 1994; Camerer, 1997). Im folgenden werden drei zentrale Präferenzen in den Praktiken der experimentellen Ökonomie – leistungskontingente Entlohnung, Betonung des Lernens, Tabuisierung von Täuschungsexperimenten – näher betrachtet, wobei hier nur Experimente zum Bayesianischen Denken (siehe Koehler, 1996) ausführlicher diskutiert werden. (Hertwig und Ortmann, 1997, diskutieren diese Präferenzen im Hinblick auf weitere Themen der psychologischen Entscheidungsforschung). Zwei Fragen sollen hier erörtert werden: Erstens, welche Gründe unterliegen den methodischen Präferenzen der Ökonomen? Zweitens, induzieren die Präferenzen Ergebnisse, die sich von denen psychologischer Experimente unterscheiden?

Leistungskontingente Bezahlung

Versuchspersonen in der experimentellen Ökonomie werden leistungskontingent bezahlt. Ausnahmen von dieser Praxis gibt es nur wenige. Roth (1995) vermutet sogar, daß „the question of actual versus hypothetical choices has become one of the fault lines that have come to distinguish experiments published in economics journals from those published in psychology journals“ (1995, S. 86). Die Grenzziehung ist sicher nicht so kategorisch, wie Roth sie darstellt,

da man zum Beispiel auch in der Forschung zum Heuristics-und-Biases-Programm gelegentlich Studien finden kann, die leistungskontingente Bezahlung verwenden oder sie zum Untersuchungsgegenstand machen (für Beispiele siehe Hertwig & Ortmann, 1997). Es ist aber sicher richtig, daß in der psychologischen Entscheidungsforschung leistungskontingente Bezahlung eher die Ausnahme ist, wohingegen die Veröffentlichung experimenteller Resultate in einem führenden Ökonomiejournal ohne leistungskontingente Bezahlung nahezu unmöglich zu sein scheint (Roth, 1995, Fußnote 30; Camerer, 1997).

Wodurch wird diese Präferenz motiviert? Einer der Ursprünge ist die sogenannte Wallis-Friedman-Kritik der Arbeiten von L. L. Thurstone (1931; siehe Roth, 1995, für eine ausführliche Darstellung). Thurstone versuchte auf der Grundlage von hypothetischen Wahlentscheidungen Indifferenz-Kurven zu identifizieren. In den hypothetischen Urteilen fragte er zum Beispiel nach der Präferenz zwischen zwei Optionen von Gebrauchsartikeln (z.B. 8 Hüte und 8 Schuhe versus 6 Hüte und 9 Schuhe). Wallis und Friedman (1942) kritisierten die Prozedur von Thurstone unter anderem mit dem Argument, daß es fraglich sei, ob eine Versuchsperson in einer artifiziellen Situation wissen könne, wie sie sich tatsächlich verhalten würde.

Leistungskontingente Bezahlung und die Art und Weise, wie sie realisiert werden soll, hat Vernon Smith (1976, 1982) in zwei Artikeln kodifiziert, die zu Klassikern der experimentellen Ökonomie geworden sind. In dem Bemühen, Laborexperimente in der Ökonomie als eine legitime Art der Erkenntnisgewinnung zu etablieren, hat er fünf Bedingungen („precepts“) des Experimentierens postuliert, deren Befolgung dem Experimentator hinreichende Kontrolle über die Präferenzen eines Individuums garantieren soll. Drei der fünf Bedingungen – *nonsatiation*, *saliency* und *dominance* – beziehen sich unmittelbar auf die Belohnungsstruktur des ökonomischen Experiments.

Die *nonsatiation*-Bedingung stellt folgende Forderung an das Belohnungsmedium v (z.B. Geld, soziale Anerkennung): Nutzen (utility) ist eine monoton steigende Funktion von v . Wenn dies erfüllt ist, dann kann man erwarten, daß eine Alternative A mit der Konsequenz v_A einer Alternative B mit der Konsequenz v_B vorgezogen wird, sofern gilt, daß $v_A > v_B$. Damit die Belohnung in einem Experiment motivationale Relevanz erlangt, muß sie *salient* sein. Das heißt zum Beispiel, daß eine Präferenz für A gegenüber B sich nur dann realisiert, wenn eine Entscheidung im Experiment mit eindeutigen Konsequenzen verbunden ist. Die *dominance*-Bedingung schließ-

lich postuliert, daß die Bezahlung so gestaltet sein soll, daß sie jegliche subjektiven Kosten, die mit der Teilnahme an den Aktivitäten im Experiment verbunden sein könnten, dominiert. Angenommen, für die Wahl zwischen A und B gelte $v_A > v_B$, dann muß diese ordinale Beziehung so gestaltet sein, daß die Ordnung auch dann erhalten bleibt, wenn subjektive Kosten δ , die mit der Wahl von Alternative A oder B verknüpft sein können, berücksichtigt werden.

Den Smithschen Bedingungen für ein valides und hinreichend kontrolliertes Experiment liegt ein *effort*-Modell des Entscheidungsprozesses (Smith & Walker, 1993) zugrunde. In dieser Vorstellung bedürfen bessere Entscheidungen (d. h. optimale Entscheidungen im Sinne des normativen Modells) mehr kognitiver Anstrengung, wobei Anstrengung als eine Aktivität mit negativem Nutzen konzeptualisiert wird. Deshalb, so das Argument, wird derjenige, der Entscheidungen trifft, einen Ausgleich herstellen zwischen dem positiven Nutzen einer optimalen Entscheidung und dem negativen Nutzen, der in Form von Anstrengungen eintritt. In dieser Konzeption sollten mit ansteigendem Belohnungsniveau für optimale Entscheidungen auch die Investitionen in Gestalt von Anstrengung zunehmen. Anders ausgedrückt: Man würde vorhersagen, daß mit zunehmender Bezahlung die Entscheidungen optimaler würden und die Wahrscheinlichkeit von zufälligen Entscheidungen abnähme. Auf Normverstöße im Sinne des Heuristics-und-Biases-Programms angewendet bedeutet dies, daß bei salienter und dominanter Bezahlung Verstöße zum Beispiel gegen probabilistische Gesetzmäßigkeiten weniger wahrscheinlich sein sollten (z. B., weil die Versuchspersonen dann motiviert wären, andere Strategien als „quick“ und „dirty“ Heuristiken anzuwenden).

Bringt leistungskontingente Bezahlung kognitive Illusionen zum Verschwinden?

Wenn dem so wäre, hätte das auch Implikationen für die experimentelle Praxis in der psychologischen Entscheidungsforschung. Ihrer Bedeutung angemessen wird die Diskussion um den Einfluß leistungskontingenter Bezahlung in der experimentellen Ökonomie mit einiger Verve geführt. Ich möchte die Diskussion anhand von zwei Personen – Richard Thaler und dem bereits erwähnten Vernon Smith – illustrieren, die entscheidend zu dieser Diskussion beigetragen haben. Thaler, ein experimenteller Ökonom, der eng mit Tversky und Kahneman zusammengearbeitet hat, kommt (zusammen mit Colin Camerer) fast allein das Verdienst zu, psychologische Entscheidungsforschung, im Sinne des Heuristics-und-Biases-Programms, in die Ökonomie importiert zu haben. Der

Import psychologischer Befunde, insbesondere jener, die die Rationalitätsannahme des ökonomischen Verhaltensmodells in Frage stellen, kann nur dann erfolgreich sein, wenn Ökonomen der Qualität der Experimente nicht mißtrauen. Mißtrauen deshalb, weil sie unter Bedingungen durchgeführt wurden, die nach Meinung vieler experimenteller Ökonomen keine hinreichende experimentelle Kontrolle garantieren.

Auf dem Hintergrund dieser Interessenlage überrascht es nicht, daß Thalers argumentative Strategie darin besteht, im Hinblick auf die Validität der Ergebnisse des Heuristics-und-Biases-Programms die Bedeutung von leistungskontingenter Bezahlung zu relativieren. Eine schöne Illustration dieser Strategie findet man in Thaler (1987a). Hier vertrat er die Meinung, daß sich Bezahlungen als irrelevant herausgestellt hätten („monetary incentives have proved to be irrelevant in many cases“, S. 120), um dann sogar noch einen Schritt weiterzugehen und den Schluß zu ziehen, daß „the violations of rationality observed tend to be somewhat *stronger* in the incentive condition“ (S. 119). Die Replik von Smith kam einige Zeit später. Bezugnehmend auf Thalers (1987b) und Tversky und Kahnemans (1987) Behauptung, wonach es wenig („little“) Evidenz für die Behauptung vieler Ökonomen gäbe, Menschen träfen bessere Entscheidungen, wenn mehr auf dem Spiel stünde, beklagt Smith (1991) deren selektiven Blick: „Strangely missing, since it is said that there is a „little“ evidence, are any of the many citations that could have been offered showing that monetary rewards matter.“ (S. 887)

In den Äußerungen von Thaler und Smith kommt zum Ausdruck, daß der Disput bereits bei der Frage beginnt, worin denn die Alternativhypothese bestehe, die es zu bestätigen oder zu verwerfen gilt. Thaler (1987b) bedient sich verschiedener Varianten: Mal spricht er davon, daß Bezahlungen Normverstöße verstärken (Thaler, 1987a; siehe für einen ähnlichen Standpunkt auch Arkes, Dawes & Christensen, 1986), dann spricht er davon, daß Bezahlung sich in vielen Fällen als irrelevant herausgestellt habe (Thaler, 1987a), um schließlich die Formulierung zu verwenden, daß „the assertion that systematic mistakes will always disappear if the stakes are large enough should be recognized for what it is – an assertion unsupported by any data“ (Thaler, 1987b, S. 96). Zu Recht, wie ich finde, kritisiert Smith (1991, S. 887) Thalers letzte Formulierung als Versuch, eine Hypothese zu etablieren – „mistakes will *always* disappear“ – derer man sich ohne große Anstrengung entledigen kann. Ernster zu nehmen ist Thalers Konklusion, Bezahlung habe sich in vielen Fällen als irrelevant herausgestellt. Also noch einmal, trifft das zu?

Mir ist kein auch nur annäherungsweise erschöpfender Überblick bekannt, der es erlauben würde, die Frage zu beantworten, ob die Befunde des Heuristics-und-Biases-Programms auch im Fall von leistungskontingenter Bezahlung stabil sind. Hertwig und Ortmann (1997, Tabelle 1) haben allerdings eine Reihe von Arbeiten aus der Psychologie zusammengestellt. Diese zeigen deutlich, daß leistungskontingente Entlohnung Urteile im Sinne des jeweiligen normativen Kriteriums durchaus verbessern kann. Leistungskontingente Bezahlung scheint das Ausmaß von „conservativism“, also der Überbewertung der *Basisraten* in Relation zu den *Likelihoods*, zu reduzieren (Phillips & Edwards, 1966), die Genauigkeit von Häufigkeitsschätzungen (Wright & Aboul-Ezz, 1988) und die Qualität von Entscheidungen in komplexen Aufgaben zu verbessern (Campbell, 1984), den Einfluß von Ankereffekten bei Wahrscheinlichkeitsurteilen (Wright & Anderson, 1989) und den Rückschaufehler zu reduzieren (Hell, Gigerenzer, Gauggel, Mall & Müller, 1988) und in Verbindung mit anderen Manipulationen das Preference-reversal-Phänomen zu eliminieren (Ordóñez, Mellers, Chang & Roberts, 1995).

Diesen Ergebnissen psychologischer Studien kann man noch eine Überlicksarbeit von Smith und Walker (1993) hinzufügen, die 31 Studien der experimentellen Ökonomie im Hinblick auf den Effekt von leistungskontingenter Bezahlung analysierten. Unter den 31 Studien fanden die Autoren mehrere, die folgendes zeigen: Je mehr Geld auf dem Spiel steht, desto größer die Konformität der Urteile mit den Vorhersagen der ökonomischen Modelle. Außerdem fanden Smith und Walker (1993), und dies ist vielleicht sogar noch wichtiger für die psychologische Entscheidungsforschung, daß „in virtually all cases rewards reduce the variance of the data around the predicted outcome“ (S. 245).

Natürlich kann man dieser Evidenz auch eine entsprechende Liste von Arbeiten gegenüberstellen, die zeigen, daß leistungskontingente Bezahlung keine normgerechteren Urteile produziert. Ein solche Liste könnte Tversky und Kahnemans Arbeiten zur Güte von relativen Häufigkeitsschätzungen (1973, Study 3 und 7), zum Preference-reversal-Phänomen (1981), zum Konjunktionsfehler (1983, S. 303) und zum Gesetz der großen Zahl (Kahneman & Tversky, 1972) umfassen. Sie könnte auch die Arbeit von Hsee und Weber (1997) zu einem Attributionsfehler bei Risiko-Präferenz-Urteilen und den Befund von Arkes et al. (1986) beinhalten, wonach sich als Folge von leistungskontingenter Bezahlung die Güte von Urteilen unter bestimmten Bedingungen verschlechtern kann.

Welche Schlüsse lassen sich aus dieser widersprüchlichen Kollektion von Befunden ziehen? Kate-

gorische Aussagen, wie zum Beispiel bei Thaler (1987a,b), sind nicht gerechtfertigt. Wir wissen wenig über den Einfluß von leistungskontingenter Bezahlung auf die Ergebnisse des Heuristics-und-Biases-Programms. Dies bedeutet aber auch, daß psychologische Entscheidungsforscher, sofern sie die wichtige Frage nach der Validität ihrer Befunde nicht ausschließlich einer anderen Disziplin überlassen wollen, dies zum Gegenstand ihres Forschungsinteresses machen müssen. Hertwig und Ortmann (1997) haben zu diesem Zweck vorgeschlagen, daß der simultane Test von Urteilsverhalten mit und ohne leistungskontingenter Bezahlung verbindlicher Bestandteil psychologischer Studien werden solle. Im Gegensatz zu den meisten gegenwärtigen Studien psychologischer Entscheidungsforscher, in denen Motivation kaum berücksichtigt wird, wäre ihre systematische Manipulation dann der Normalfall. Motivation zu vernachlässigen wäre die Ausnahme, die es zu begründen gilt.

Es gibt aber auch noch eine ganz pragmatische Erwägung, warum Experimentatoren in der Psychologie ausgezeichnete Gründe hätten, leistungskontingente Bezahlung zu konventionalisieren. Selbst Thaler (1987a, S. 119), der die Bedeutung leistungskontingenter Bezahlung relativiert, zieht nicht in Zweifel, daß es einen sekundären Effekt von Bezahlung gibt, der Experimentieren leichter machen kann: Leistungskontingente Bezahlung kann die Fehlervarianz reduzieren (Smith & Walker, 1993). Ein konkretes Beispiel: In seinen Studien zur Mißachtung von Basisraten fand Grether (1992), daß „the rate of incoherent or nonsense responses was three times higher for those subjects lacking financial incentives“ (S. 54).⁵ Dieser Befund ist unabhängig von der Frage, ob Menschen sich im Sinne der ökonomischen Modelle, der Wahrscheinlichkeitstheorie, der Logik etc. rational verhalten oder nicht. Sowohl die Suche nach Evidenz für die deskriptive Validität normativer Modelle wie auch für Evidenz zugunsten alternativer Modelle (z. B. Heuristiken) wäre erleichtert, wenn die Fehlervarianz geringer wäre.

Die Konventionalisierung leistungskontingenter Bezahlung allein reicht nicht aus. Es ist darüber hinaus notwendig, sie konzeptuell besser zu verstehen. Dazu gehört zum Beispiel ein besseres Verständnis des „flat maximum“-Problems: Obwohl die Explikation der Kosten und Nutzen von Verhalten (z. B. Urteil, Entscheidung etc.) eine effektive Instruktions-

⁵ Der Prozentsatz inkohärenter Antworten betrug 4% (mit finanzieller Entlohnung) versus 12% (ohne Entlohnung); der Prozentsatz von Versuchspersonen, die mindestens eine inkohärente Antworten gab, betrug 31% (15 von 48; mit Entlohnung) versus 65% (32 von 49; ohne Entlohnung).

methode ist (siehe Edwards, 1961), schafft ein *flat maximum* das Problem, daß non-optimales Verhalten oft nicht genügend bestraft wird (von Winterfeldt & Edwards, 1982). Die Konsequenz ist, daß, selbst wenn man sich nicht optimal verhält, dies nicht wesentlich den erwarteten Nutzen (im Vergleich zu optimalem Verhalten) reduziert.

Natürlich ist Bezahlung nicht gleich Bezahlung. Smith (1982) hat einige der Bedingungen für eine effiziente Bezahlung definiert: Leistungskontingente Bezahlung sollte salient und dominant sein. Harrison (1994) hat an den wenigen Experimenten von Tversky und Kahneman, in denen diese leistungskontingente bezahlten, kritisiert, daß diese der *dominance*-Bedingung nicht genügen. Es geht weiterhin darum, besser zu verstehen, in welchen Domänen leistungskontingente Bezahlung sinnvoll ist (eine simple, notwendige Bedingung für solche Domänen ist, daß es eine objektiv korrekte Reaktion gibt, die dann entsprechend verstärkt werden kann⁶), und in welchen Domänen Bezahlung das Verhalten zerstört, welches es zu untersuchen gilt (z. B. Spielverhalten eines Kindes, siehe hierzu den Überblick von Eisenberger & Cameron, 1996, zum *overjustification effect*).

Schließlich gilt es auch, die Kosten-Nutzen-Erwägungen, die mit der Bezahlung, aber auch Nicht-Bezahlung einhergehen, transparent zu machen. Leistungskontingente Bezahlung erfordert in der Regel zwar umfangreichere Instruktionen, aber gleichzeitig zwingen diese Instruktionen Experimentatoren zu Vollständigkeit und Klarheit gegenüber den Versuchspersonen. Leistungskontingente Bezahlung macht Experimente teurer, aber sie garantiert möglicherweise die größeren und stabileren Effekte (via kleinerer Fehlervarianz). Dies ist um so bedeutsamer, als die Psychologie, wie Geoffrey Loftus es sieht, sich noch immer im Zustand einer Landschaft „dotted with dense stands of conflicting data that strangle theoretical advances at their roots“ (in Bower, 1997, S. 356) befindet. Natürlich ist die Empfehlung leistungskontingenter Anreize keineswegs neu in der Psychologie, sondern anscheinend nur in Vergessenheit geraten: Für psychologische Forschung zum Beispiel im Rahmen des behavioristischen Paradigmas oder innerhalb des konzeptuellen Rahmens der Signalentdeckungstheorie waren sie ein integraler Bestandteil des Experimentierens.

⁶ Vielleicht versteckt sich hier einer der Gründe, warum Studien zum Einfluß leistungskontingenter Bezahlung auf die Existenz kognitiver Illusionen widersprüchliche Resultate hervorgebracht haben: Die untersuchten Probleme sind gelegentlich so formuliert, daß mehr als ein normatives Kriterium Anwendung finden könnte (Birnbaum, 1983; Gigerenzer, 1996).

Wiederholung experimenteller Versuchsdurchgänge

Ökonomen sind häufig an Gleichgewichtszuständen und dem Lernprozeß, der zu diesen führt, interessiert. Dieses Interesse diktiert eine weitere methodische Präferenz der experimentellen Ökonomie, nämlich die Wiederholung von experimentellen Versuchsdurchgängen (im folgenden Trials genannt). Thaler (1987a, S. 119) räumt ein, daß – im Gegensatz zu dieser Präferenz – die Studien des Heuristics-und-Biases-Programms den Versuchspersonen wenig Chancen zum Lernen lassen. Damit ähneln die Fronten denen, die man bei der Diskussion um die Bedeutung von leistungskontingenter Bezahlung beobachten konnte – und die Argumente der Kontrahenten werden vorhersagbar.

Warum legen Ökonomen Wert auf die wiederholte Darbietung experimenteller Trials? Diese Präferenz folgt unmittelbar aus dem Konzept des Gleichgewichtszustandes, das für ökonomische Theorien zentrale Bedeutung hat. Camerer (1997) hat die Bedeutung wie folgt zusammengefaßt:

Economic theories are usually asserted to apply to equilibrium behavior, after initial ‘confusion’ disappears and (unspecified) equilibrating forces have had time to operate. So E’s generally conduct an experiment for several repeated trials under ‘stationary replication’ – destroying old goods and refreshing endowments, in order to recreate the decision situation in exactly the same way each time (except for anything that was learned from the previous experience ...). Then special attention is paid to the last periods of the experiment (e.g., often the last half of the trials are used as data to test an hypothesis) or to the change in behavior across trials. Rarely is rejection of a theory using first-round data given much significance. (S. 318–319)

Camerer betont eine Form des Lernens: das allmähliche Verstehen der experimentellen Umgebung (Prozedur, Instruktionen etc.) durch eine zu Beginn desorientierte Versuchsperson. Ökonomen ermöglichen ihren Versuchspersonen diese Art des Lernens mittels *stationary replication*. In Entscheidungssituationen, die Interaktionen mit anderen Personen beinhalten, bedeutet *stationary replication*, daß Trials wiederholt werden, diese Wiederholungen aber jeweils mit einem neuen Gegenüber stattfinden. Der Vorteil besteht darin, daß eine Versuchsperson die experimentelle Situation erfahren kann, ohne gleichzeitig mit der Frage beschäftigt zu sein, welche Strategie die beste sei, wenn man mit der gleichen Person mehr als einmal interagiert. Dieser letzte Aspekt beschreibt eine weitere Form von Lernen: das Verstehen der strategischen Aspekte einer Interaktion, und wie diese sich in Abhängigkeit davon, ob man mit der gleichen Person noch einmal zusammentreffen kann, ändern.

Aus spieltheoretischer Perspektive macht es einen großen Unterschied, ob ein Spiel, zum Beispiel das Prisoner's-Dilemma-Spiel, einmal gespielt wird (*single shot*), endlich häufig oder unbestimmt (*indefinite*) häufig wiederholt wird. Sowohl für das einmalige wie auch für das endlich häufig wiederholte Spiel besteht die optimale Strategie in unkooperativem Verhalten, wohingegen in einem unbestimmt häufig wiederholten Spiel kooperatives Verhalten optimal ist. Um dies zu verstehen, muß der Spieler aber lernen, strategische Abwägungen zu treffen, die auch das Verhalten und die Agenda der anderen Spieler berücksichtigen. Dieses Verständnis kann man nicht voraussetzen, sondern der Experimentator muß die Entscheidungssituation so gestalten, daß die Versuchsperson diese Kompetenz erwerben kann.

Zusammengefaßt gibt es also mindestens zwei Motive für die methodische Präferenz, experimentelle Trials zu wiederholen: den Versuchspersonen die Möglichkeit zu geben, erstens die Aspekte der experimentellen Situation (z.B. Prozedur, Instruktion) und zweitens die strategischen Aspekte der Interaktionen zu verstehen. Ist es aber nicht auch in der psychologischen Entscheidungsforschung etablierter Standard, experimentelle Trials zu wiederholen? In der Forschung zum „Overconfidence“-Phänomen zum Beispiel (z.B. Lichtenstein, Fischhoff & Phillips, 1982; Gigerenzer, Hoffrage & Kleinbölting, 1991) werden Versuchspersonen gebeten, nicht nur einzelne Fragen, sondern ganze Listen von Fragen zu beantworten (z.B. der Art „Welche Stadt hat mehr Einwohner: Chicago oder Miami?“). Aus der Sicht der experimentellen Ökonomen handelt es sich hier aber nicht um wiederholte Trials, weil in der Regel eine notwendige Bedingung für Lernen, nämlich Rückmeldung über die Performanz, nicht erfüllt ist.

In Kombination mit expliziter oder impliziter (in Form leistungskontingenter Bezahlung) Rückmeldung ist die Wiederholung von experimentellen Trials also bedeutsam in der Ökonomie, wohingegen in der psychologischen Entscheidungsforschung in erster Linie *single-shot*-Experimente durchgeführt werden. Natürlich überrascht es daher kaum, wenn auch diese unterschiedlichen Präferenzen und ihre Implikationen für die Validität der Resultate Gegenstand heftiger Diskussion sind. Die Frontstellung ist die alte. Auf der einen Seite diejenigen, die psychologische Resultate in die Ökonomie importieren wollen, und auf der anderen Seite viele experimentelle Ökonomen, die deren Validität anzweifeln. Thaler (1987a) steht stellvertretend für die erste Gruppe, und er argumentierte, für *single-shot*-Experimente spräche, daß „many decisions are made infrequently We marry, choose careers, and take jobs a few times at most.“ (S. 122) Außerdem zeigte er sich überzeugt, daß „there is every reason to believe that an ini-

tial response in the laboratory will most likely be the one a subject will make in a similar real-life problem“ (S. 121). Ähnlich wie Thaler argumentierten zum Beispiel auch Tversky und Kahneman (1987): „... most important decisions are unique and therefore provide little opportunity for learning“ (S. 90).

Vermutlich ist die Frage ziemlich müßig, ob die meisten wichtigen Entscheidungen einzigartig sind oder nicht, denn was konstituiert eine einzigartige Situation? Hat nicht jede Entscheidung Merkmale mit anderen gemeinsam? Konstruktiver scheint die Frage, ob es Evidenz dafür gibt, daß die Ergebnisse des Heuristics-und-Biases-Programms anders aussähen, wenn mehr Augenmerk auf Lernprozesse via Wiederholung experimenteller Trials und Feedback gelegt würde. Es ist mir zwar kein Überblick bekannt, der es erlaubt, diese Frage abschließend zu beantworten, aber es gibt eine Reihe von Studien, die demonstrieren, daß die Wiederholung experimenteller Trials großen Einfluß auf die Existenz kognitiver Täuschungen haben kann. Ich werde im folgenden drei kurz beschreiben, die das Phänomen Mißachtung von Basisraten zum Gegenstand haben. Mit diesen Studien soll nichts über die Prävalenz dieses Lerneffekts ausgesagt sein, sehr wohl aber etwas über die potentielle Bedeutung von Lernen.

Bringt die Wiederholung experimenteller Trials kognitive Illusionen zum Verschwinden?

Harrison (1994) hat den Einfluß von Lernen auf die Fähigkeit seiner Versuchspersonen, in Übereinstimmung mit dem Bayes-Theorem entscheiden zu können, untersucht. Sein computersimuliertes Szenario bestand aus drei Urnen. Urne *C* enthielt 6 Kugeln, die fortlaufend nummeriert waren. Eine Ziehung aus ihr determinierte, ob eine weitere Kugel aus Urne *A* oder *B* gezogen wurde. Zum Beispiel konnte die Regel lauten: „Wenn die Zahlen 1 bis 4 gezogen werden, wird Urne *A* gewählt, sonst *B*.“ Den Versuchspersonen wurde zwar diese Regel, aber nicht der Ausgang der ersten Ziehung mitgeteilt. Weiterhin kannten die Versuchspersonen die Komposition der Urnen *A* (4 rote Kugeln und 2 schwarze Kugeln) und *B* (3 rote und 3 schwarze Kugeln). Die Versuchspersonen beobachteten die Ziehung von 6 Kugeln (mit Zurücklegen) aus entweder Urne *A* oder *B* (welche es war, wurde den Versuchspersonen nicht mitgeteilt). Anschließend wurden sie gebeten, entweder auf Urne *A* oder *B* als dem wahrscheinlichen Ursprung der sechs Ziehungen zu wetten.

Den Versuchspersonen standen also zwei Informationen zur Verfügung: die Basisraten (in Form der „Wenn, dann“-Regel) und der Likelihood-Quotient ($p(\text{rot}|Urne A) / p(\text{rot}|Urne B)$). Das Theorem von

Bayes gibt die Antwort darauf, wie diese integriert werden sollen, um die a-posteriori-Wahrscheinlichkeit der Urne *A* (bzw. *B*) bestimmen zu können. Harrisons Experiment bestand aus zwei Blöcken mit jeweils zehn Ziehungs- bzw. Urteilsdurchgängen. Am Ende jeder Entscheidung erhielten die Versuchspersonen Rückmeldung darüber, ob sie sich jeweils für die richtige (im Sinne des Bayes-Theorems) Urne entschieden hatten, und jeweils am Ende eines Blocks wurden zwei Gruppen von Versuchspersonen leistungskontingent bezahlt (entweder auf der Grundlage einer zufällig ausgewählten Entscheidung oder aller Entscheidungen), und eine Gruppe erhielt einen leistungsunabhängigen Festbetrag. Der zweite Block von zehn Entscheidungen ist eine Wiederholung des ersten Blocks mit neuen Zufallsziehungen – allerdings hatten die Versuchspersonen die Möglichkeit, im ersten Block zu lernen.

In seiner Analyse verglich Harrison (1994, Tabelle 2) die Anzahl korrekter Entscheidungen (im Sinne des Bayes-Theorems) in den Fällen, in denen die Vorhersagen der Repräsentativitäts-Heuristik mit denen des Bayes-Theorems konsistent sind, und in den Fällen, in denen die Vorhersagen inkonsistent sind. Die Auswertung hatte folgende Logik: Wäre der Prozentsatz korrekter Entscheidungen größer für die konsistenten Fälle als für die inkonsistenten Fälle, dann spräche dies dafür, daß die Urteilenden die Heuristik angewendet haben. War dies der Fall? Harrison beobachtete zwei Effekte. Erstens fand er, daß gemittelt über die beiden Gruppen, in denen leistungskontingent bezahlt wurde, der Prozentsatz korrekter Entscheidungen entweder nur geringfügig größer war für die konsistenten Fälle als für die inkonsistenten Fälle (6 Prozentpunkte in Block 1) oder sogar identisch war (0 Prozentpunkte in Block 2). Im Gegensatz dazu fand er, daß es in der Gruppe, in der nicht leistungskontingent bezahlt wurde, deutlich mehr korrekte Entscheidungen für die konsistenten Fälle gab (30 Prozentpunkte in Block 1). Laut Harrison (1994) ist dies Evidenz dafür, daß „the Heuristic strongly influences the propensity of unmotivated and inexperienced subjects to make correct decision,“ (S. 249).

Zweitens fand er einen Lerneffekt. Der kam darin zum Ausdruck, daß der 30-Prozentpunkte-Unterschied im ersten Block für jene Versuchspersonen, die nicht leistungskontingent bezahlt wurden, im zweiten Block fast vollständig verschwand (auf nur noch 4 Prozentpunkte). Harrison (1994) kam zu der Schlußfolgerung, daß „when these unmotivated and experienced subjects tackle the same class of problems the Heuristic has no noticeable influence at all“ (S. 250).

Harrisons (1994) experimentelles Design war eng an das von Grether (1980) angelehnt. Im Unterschied

zu Harrison fand Grether (1980, 1992), daß die Urteile der Versuchspersonen dazu tendierten, mit den vorhergesagten Urteilen der Repräsentativitäts-Heuristik übereinzustimmen. Grether (1980) gelangte weiterhin zu der Schlußfolgerung, daß „monetary incentives do not appear to affect the behavior of inexperienced subjects This is not true, however, for the experienced subjects“ (Grether, 1980, S. 552; Hervorhebungen von Grether). Die letzte Beobachtung, die nahelegt, daß die Vorhersagen von finanziell motivierten und erfahrenen Versuchspersonen eher mit dem Bayes-Theorem konsistent sind, ist insofern bemerkenswert, als Grether (1980) im Unterschied zu Harrison (1994) eine ungebräuchliche Definition von Erfahrung verwendete. Danach galt jemand bereits dann als erfahren, wenn die Person schon einmal die gleiche Kombination von Likelihood-Quotient und Basisrate beurteilt hatte (normalerweise gilt eine Person dann als erfahren, wenn sie Feedback über ihre Leistung erhalten hat). Auf der Grundlage dieser Definition schloß Grether, daß „there was no feedback during the course of the experiment, so any ‚learning‘ that might have taken place results from familiarity with a problem or from further reflections on it“ (S. 553).⁷

Camerer (1987, 1990) hat den Einfluß von Lernen und Erfahrung auf Bayesianisches Urteilen in einem Marktkontext („asset market“) untersucht. Die Logik des Bayesianischen Marktexperimentes wurde bereits erwähnt: Der Markt ist so strukturiert, daß der Gewinn aus Aktienkäufen und -verkäufen dadurch maximiert werden kann, daß man die Eintretenswahrscheinlichkeit von zwei Ereignissen möglichst genau einschätzt. Die Marktteilnehmer (Versuchspersonen) schätzten zuerst die Wahrscheinlichkeit der Ereignisse ein und handelten dann Aktien in einem Markt, in dem die Dividenden der Aktien davon abhingen, welches der beiden Ereignisse tatsächlich eintrat. Das Marktgeschehen bestand aus mehreren Perioden, und nach jeder Marktperiode erhielten die Versuchspersonen qua ihres Gewinnes Rückmeldung über ihre Performanz. Camerer hat dieses Szenario in 15 Experimenten untersucht. Hier sollen lediglich seine Schlußfolgerungen berichtet werden:

In 15 experiments, trading prices and allocations showed consistent, but small, representativeness bias where we predicted it to occur. For the samples in which we expected less representativeness, prices were very close to Bayesian expected values. With experienced subjects, price biases were substantially eroded, but did not disappear; and allo-

⁷ In diesem Zusammenhang ist auch die folgende Beobachtung Grethers (1980) erwähnenswert. Ähnlich wie leistungskontingente Bezahlung kann auch Erfahrung die Fehlervarianz reduzieren: „In all but one case ... the experienced subjects showed less random or ‚more precise‘ behavior than the inexperienced subjects.“ (S. 553)

cation data still suggested representativeness.⁸ (Camerer, 1990, S. 168)

In einer weiteren Reihe von 7 Experimenten untersuchte Camerer (1990), ob die Repräsentativitäts-Heuristik den fälschlichen Glauben, Insiderwissen zu besitzen („information mirages“), hervorrufen könnte. Hier gelangte Camerer (1990) zu der Schlussfolgerung: „Representativeness can cause mirages before subjects learn, but these data cast doubt on the theory that representativeness causes chronic stock market volatility among experienced traders in natural markets.“ (Camerer, S. 165, 1990; Hervorhebung von Camerer)

Die drei Studien – Harrison (1994), Grether (1980) und Camerer (1990) – dokumentieren den Einfluß von Lernprozessen auf einen Effekt, Basisraten-Mißachtung, der üblicherweise als äußerst stabil erachtet wird. In diesen Studien deutet sich die Möglichkeit an, daß Lernen eine ähnlich effiziente Methode ist, Normverstöße zu reduzieren oder zum Verschwinden zu bringen, wie der Wechsel des Repräsentationsformats von Wahrscheinlichkeiten zu Häufigkeiten (siehe Fiedler, 1988, Gigerenzer & Hoffrage, 1995).⁹ Diese Möglichkeit zu eruieren, sollte eine Aufgabe nicht nur der experimentellen Ökonomie, sondern auch der Psychologie sein. Hertwig und Ortmann (1997), die den Einfluß von Lernprozessen auch auf andere kognitive Illusionen (z. B. Preference-reversal-Phänomen) und im spieltheoretischen Kontext diskutieren, schlagen vor, daß Studien der psychologischen Entscheidungsforschung explizit eine Lernperspektive einbeziehen sollten.

⁸ Camerer (1990) berichtete, daß er in den 15 Experimenten nur einen kleinen, „representativeness bias“ fand: die absolute mittlere Abweichung von der Bayesianischen Lösung betrug für erfahrene Versuchspersonen nur .07 und für unerfahrene .12 (siehe Camerer, 1990, Table 5, letzte Spalte, für 1-red und 2-red samples).

⁹ Auch in der Psychologie haben einige Studien den Einfluß von Feedback und Lernprozessen auf die Mißachtung von Basisraten exploriert. Es ist aber vermutlich kein Zufall, daß diese Studien eher im Bereich der Kategorisierungsliteratur (z. B. Gluck & Bower, 1988; Estes, Campbell, Hatsopoulos & Hurwitz, 1989; Kruschke, 1996) und nur selten im Bereich der psychologischen Entscheidungsforschung (z. B. Lindeman, van den Brink & Hoogstraten, 1988) angesiedelt sind, und manchmal von erklärten Behavioristen (Goodie & Fantino, 1995, 1997) durchgeführt wurden. Das Bild, das man von diesen Arbeiten erhält, ist (noch) relativ widersprüchlich: Zum Beispiel scheint es so zu sein, daß Feedback und Wiederholung die Urteile (gemessen am Bayes-Theorem) deutlich verbessern (z. B. Lindeman et al., 1988; Estes et al., 1989; siehe aber auch Goodie & Fantino, 1995), diese Leistungssteigerung sich aber nicht notwendigerweise auf andere Aufgaben bzw. Test-Sets überträgt (z. B. Lindeman et al., 1988, Gluck & Bower, 1988; Estes et al., 1989).

Dieser Vorschlag ist wiederum nicht neu. In der psychologischen Entscheidungsforschung wurde wiederholt (vergeblich) gefordert, Lernprozessen (und Feedback aus der Umgebung) mehr Aufmerksamkeit zu schenken (z. B. Winkler & Murphy, 1973). Vielleicht am deutlichsten war Hogarth (1981), der spekulierte, daß viele Urteilsfehler möglicherweise das Produkt von solchen Experimenten sind, die kontinuierliche und feedbackabhängige Urteilsprozesse mittels diskreter Ereignisse (z. B. einmaligen Urteilen) untersuchen. Im einem ganz anderen Bereich psychologischer Forschung, der Exploration von Altersprozessen, wird eine Methodologie (*Testing-the-limits*, z. B. Kliegl, Smith & Baltes, 1989) erfolgreich eingesetzt, deren Ziel die schrittweise Ermittlung des maximalen Leistungspotentials ist. Im Gegensatz dazu sind Entscheidungsforscher in der Psychologie, in den Worten von Mellers, der früheren Präsidentin der „Society for Judgment and Decision Making“, „well known for setting traps and taking delight at human failure“ (S. 3). Woher rührt diese Fixation auf Normverstöße und warum interessieren wir uns so wenig für das Leistungspotential unserer Versuchspersonen?

Keine Täuschungsexperimente

In ihren verschiedenen Erklärungen des Preference-reversal-Phänomens haben Grether und Plott (1979) auch die Möglichkeit erwogen, daß die psychologischen Experimentatoren selbst die Ursache dieses Phänomens sind. Grether und Plott machen die Gründe für ihr Mißtrauen ziemlich klar: „Subjects nearly always speculate about the purpose of experiments and psychologists have the reputation for deceiving subjects.“ (S. 629) Im Gegensatz zur Psychologie wird die Täuschung von Versuchspersonen in der experimentellen Ökonomie als Tabu behandelt. Die Gründe erklären Davis und Holt (1993) wie folgt:

The researcher should ... be careful to avoid deceiving participants. Most economists are very concerned about developing and maintaining a reputation among the student population for honesty in order to ensure that subject actions are motivated by the induced monetary rewards rather than by psychological reactions to suspected manipulation. Subjects may suspect deception if it is present. Moreover, even if subjects fail to detect deception within a session, it may jeopardize future experiments if the subjects ever find out that they were deceived and report this information to their friends. (S. 23–24)

Was diese Argumentation von manch anderer unterscheidet, die sich in der Psychologie gegen den Gebrauch von Täuschung wendet (z. B. Baumrind, 1964, 1985), ist, daß keine ethischen, sondern rein pragmatische Erwägungen angeführt werden. Öko-

nomen befürchten Rufschädigung und Verlust ihrer Glaubwürdigkeit. Bis zum Verlust experimenteller Kontrolle wäre es dann nur ein kleiner Schritt: Hegt die Versuchsperson den Verdacht, daß das experimentelle Szenario etwas anderes vorgaukelt als tatsächlich untersucht wird, dann reagiert die Versuchsperson auf die vermutete Bedeutung der Situation und nicht auf die, die der Experimentator intendiert.

Für viele experimentelle Ökonomen ist diese Argumentation ein wichtiger Grund, den Ergebnissen psychologischer Experimente mit einer Portion Mißtrauen entgegenzutreten (siehe für eine Ausnahme Bonetti, in Druck). Ledyard (1995) schrieb zum Beispiel: „It is believed by many undergraduates that psychologists are intentionally deceptive in most experiments. If undergraduates believe the same about economists, we have lost control.“ (S. 134) Ähnlich äußerte sich auch Hey (1991), der meinte, daß es von äußerster Wichtigkeit sei, daß

economics experiments actually do what they say they do and that subjects believe this. I would not like to see experiments in economics degenerate to the state witnessed in some areas of experimental psychology where it is common knowledge that the experimenters say one thing and do another. This would be very harmful to experimental economics. (S. 119)

Könnten psychologische Entscheidungsforscher diese Überzeugung nicht mit dem Argument zu den Akten legen, daß die Verwendung von Täuschung in erster Linie in der Sozialpsychologie vorkäme (siehe Analysen in Gross & Fleming, 1982; Adair, Dushenko & Lindsay, 1985; Nicks, Korn & Mainieri, 1997)? Dem würden experimentelle Ökonomen möglicherweise zwei Dinge entgegenhalten: Erstens, Rufschädigung und experimenteller Kontrollverlust wären nur dann kein Problem, wenn man die nur schwer zu rechtfertigende Annahme macht, daß Versuchspersonen zwischen Experimenten in verschiedenen psychologischen Labors unterscheiden würden. Zweitens, stimmt es denn, daß Versuchspersonen nur in sozialpsychologischen Experimenten getäuscht werden? Um das zu beurteilen, wäre eine Definition von Täuschung hilfreich. Rosenthal und Rosnow (1991, S. 237–238) haben eine pragmatische Unterscheidung zwischen *intentional deception by omission* (d. h. passives Täuschen) und *intentional deception by commission* (aktives Täuschen) vorgeschlagen. Bei aktiver, mit Absicht herbeigeführter Täuschung durch die Informationen, die den Versuchspersonen zur Verfügung gestellt werden, ist der Fall eindeutig. Anders verhält es sich bei passiver Täuschung mittels Auslassung von Information – hier bewegt man sich in einem Graubereich.¹⁰ Der

¹⁰ John Hey, der die erste ausführliche Darstellung der experimentellen Ökonomie schrieb, vertritt einen sehr pragmatischen Standpunkt im Hinblick auf *intentional de-*

Diskussion um den Gebrauch von Täuschung kann sich die psychologische Entscheidungsforschung aber kaum mit dem Hinweis auf diesen Graubereich entziehen, da sie auch Gebrauch von aktiver Täuschung macht. Das klassische *engineer-lawyer*-Problem, mit dem Kahneman und Tversky (1973) die Mißachtung von Basisraten demonstrierten, ist dafür ein gutes Beispiel; auch deshalb, weil es die unterschiedliche Einstellung zur Verwendung von Täuschung in der Psychologie und Ökonomie illustriert.

In diesem Problem (siehe Fußnote 2) sind mehrere Behauptungen falsch: Falsch ist, daß Psychologen Interviews und Persönlichkeitstests durchgeführt hätten, auf deren Grundlage dann Persönlichkeitsprofile erstellt wurden. Falsch ist, daß die zu beurteilenden Profile zufällig gezogen worden wären, und ebenfalls falsch ist, daß Experten die Profile beurteilt hätten, ihre Urteile sehr genau gewesen wären und die Versuchspersonen danach bezahlt würden, wie genau sie die Urteile der Experten trafen. Im Unterschied zu Kahneman und Tversky legte Grether (1980) höchsten Wert auf ein authentisches Szenario und bemühte sich, dem Mißtrauen der Versuchspersonen vorzubeugen. Zum einen tat er dies, indem er Zufallsziehungen nicht nur per Instruktion ankündigte, sondern sie vor den Augen der Versuchspersonen demonstrierte. Zum anderen machte er Gebrauch von einem *Monitor*. Ein Monitor ist eine Person, die Versuchspersonen unter sich auswählen, deren alleinige Aufgabe darin besteht, das experimentelle Setting und die Zufallsziehungen zu beobachten. Das folgende ist ein Auszug aus Grethers (1980) Methodenbeschreibung:

The subjects elected one of themselves to be a monitor. The monitor was allowed to inspect all equipment, and more importantly, to observe all procedures during this experiment. It was stated in the instruction that the monitor 'should check the truthfulness of what the experimenter says, but other than that may not communicate any information to you in any way. If the monitor communicates any other information, he or she will be asked to leave without payment.' It appeared that this solved the credibility problem. (S. 541)

Die Glaubwürdigkeit des experimentellen Szenarios und letztendlich des Experimentators wird in der Ökonomie weit mehr als in der Psychologie als ein kollektives Gut erachtet, in das zu investieren sich lohnt. Sind diese Investitionen berechtigt? Oder an-

ception by omission. Er schreibt, daß *ill-defined* Experimente, das heißt Experimente, in denen nicht alle Informationen der Versuchsperson von Beginn an mitgeteilt werden, eine wichtige Art von Experimenten in der Ökonomie darstellen (in Druck). Er glaubt, daß „there is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not.“ (Hey, in Druck, S. 1)

ders gefragt: Sind die Konsequenzen der Rufschädigung durch den Gebrauch von Täuschung so dramatisch, wie Ökonomen annehmen?

Verursachen Täuschungsexperimente Kontrollverlust?

Die Psychologie hat sich in einer Reihe von empirischen Studien mit der Frage auseinandergesetzt, was die möglichen langfristigen Konsequenzen des Täuschens sein könnten. Die Schlußfolgerungen daraus sind widersprüchlich: Manche diagnostizieren den Kontrollverlust, den experimentelle Ökonomen befürchten, andere sehen dafür wenig Anlaß. Im folgenden soll eine einfache Unterscheidung helfen, die Ursache dieser widersprüchlichen Befunde besser zu verstehen.

Man kann verschiedene Methoden wählen, um die Einstellung von Versuchspersonen gegenüber psychologischen Experimenten zu destillieren. Eine Reihe von Autoren haben dies in Gestalt von (Einstellungs-)Fragebögen getan und dabei unter anderem folgendes gefunden: Versuchspersonen, die während der Experimente getäuscht wurden, berichteten nach der Aufklärung, daß sie die Experimente genossen und daß sie mehr von diesen Experimenten profitierten als von jenen, die keine Täuschung verwendeten (Smith & Richardson, 1983). Es scheint auch der Fall zu sein, daß Versuchspersonen auf Nachfrage jene Argumente unterstützen, die psychologische Experimentatoren anführen, um Täuschungsexperimente zu legitimieren (Sharpe, Adair & Roese, 1992), und daß sie im allgemeinen weniger negativ auf ethisch fragwürdige Prozeduren reagieren als Mitglieder von Ethikkommissionen, die über die entsprechenden Studien zu entscheiden haben (Smith & Berard, 1982). Sie scheinen grundsätzlich kaum Einwände dagegen zu haben, von Experimentatoren hinters Licht geführt zu werden (Christensen, 1988).

Anstatt Einstellungen zu erfragen, kann man auch Verhalten beobachten. In einigen Studien wurde dieser Zugang gewählt und dabei wurde folgendes gefunden: Versuchspersonen halten sich keineswegs immer an ein *principle of cooperativeness* (Grice, 1975), das alltägliche Konversationen auszuzeichnen scheint. Zum Beispiel teilen Versuchspersonen dem Experimentator keineswegs mit, wenn sie während des Experiments (unzulässigerweise) Zugang zu Information hatten, die ihre Reaktionen beeinflussten (Newberry, 1973; Taylor & Shepperd, 1996). Weiterhin gibt es Berichte über Vorfälle in Experimenten, die andeuten, daß Versuchspersonen dem Szenario mißtrauten und dies ihr Verhalten in der Situation beeinflusste (z. B. MacCoun & Kerr, 1987). Damit

einher geht die Erwartung, nach einer Täuschung bei weiteren psychologischen Experimenten erneut getäuscht zu werden (Krupat & Garonzik, 1994).

Also noch einmal: Gibt es den befürchteten Kontrollverlust? Wenn man lediglich Einstellungen erfragt, gibt es scheinbar wenig Grund zur Sorge. Beobachtet man allerdings tatsächliches Verhalten, dann gibt es deutliche Hinweise auf den Kontrollverlust, den Ökonomen fürchten (für weitere Studien, die sich in dieses Raster einordnen lassen, siehe Hertwig & Ortmann, 1997).

Ortmann und Hertwig (1997) schlugen vor, die Interaktion zwischen Experimentator und Versuchsperson spieltheoretisch zu betrachten. Sie besitze die Verstärkerstruktur eines Prisoner's-Dilemma-Spiels. Das heißt, es steht der Versuchsperson und dem Experimentator frei, sich kooperativ oder nicht-kooperativ zu verhalten. Nun kann man aus theoretischen und empirischen Gründen erwarten, daß andauernde Nicht-Kooperation in Gestalt von Täuschung in (unbestimmt häufig) *wiederholten* Prisoner's-Dilemma-Spielen das (pareto-optimale) Gleichgewicht zerstört, in welchem sowohl der Experimentator wie auch die Versuchsperson die kooperative Verhaltensalternative wählen würden. Faktisch handelt es sich bei Experimenten mit menschlichen Versuchspersonen um *wiederholte* Spiele in dem Sinne, daß Versuchspersonen mit bestimmten Erwartungen an das psychologische Experiment – zum Beispiel der Erwartung, getäuscht zu werden – das Labor betreten. Sollte diese Analyse korrekt sein, dann folgt daraus die klare Empfehlung, keine Täuschungsexperimente (im Sinne von „intentional deception by commission“) durchzuführen. In der Tat machen Ortmann und Hertwig (1997) diesen Vorschlag.

Sicher würde dies bedeuten, bestimmte Phänomene in Zukunft anders zu untersuchen. Aber daß man zum Beispiel kognitiven Illusionen und Heuristiken auch ohne Täuschung explorieren kann, illustrieren die Studien von Grether (1980), Camerer (1990) und Harrison (1994).

Diskussion

Fragen danach, ob und wie eine Versuchsperson motiviert werden soll, ob sie die Möglichkeit haben soll zu lernen, und ob sie Einsicht in den Gegenstand der Untersuchung haben soll (Täuschen), sind Fragen, die jeder Experimentator beantworten muß. Die jeweiligen Antworten definieren die experimentelle Situation und reflektieren die Konzeptualisierung der Versuchsperson. Im Unterschied zu Psychologen gestalten Ökonomen die experimentelle Situation aktiv.

Sie tun dies, indem sie Anreize schaffen, die mögliche andere Motive der Versuchsperson dominieren. Sie kanalisieren die „Sinnsuche“ der Versuchsperson, indem sie in ihre Glaubwürdigkeit als Experimentatoren investieren und die Versuchsperson nicht aktiv täuschen. Sie geben der Versuchsperson die Gelegenheit zum Lernen, das heißt zur Überwindung ihrer anfänglichen Desorientierung und zum Verstehen der strategischen Implikationen des experimentellen Szenarios.

Psychologen tolerieren eine ganz andere experimentelle Situation. Pointiert formuliert geben sie sich wenig Mühe, die Motivation der Versuchspersonen zu kontrollieren, räumen den Versuchspersonen kaum Gelegenheit zum Lernen ein, und kümmern sich wenig um die eigene Reputation und damit um die Erwartungen der Versuchspersonen. Wenn diese Zustandsbeschreibung annähernd zutrifft, dann könnte die Adaption der methodischen Präferenzen der Ökonomen einen Weg eröffnen, wie die Psychologie mehr Kontrolle über wichtige Parameter der experimentellen Situation – Motivation, Erwartungen und Situationsverständnis der Versuchspersonen – erlangen könnte (Hertwig & Ortmann, 1997). Dies wäre in Anbetracht der häufigen Klagen in der Psychologie über hohe Fehlervarianz, kleine Effektgrößen und Mangel an Replizierbarkeit kein unbedeutender Fortschritt.

Bounded Rationality: Eine Programm für die Psychologie und Ökonomie

Die Grenzen der Disziplinen Psychologie und experimentelle Ökonomie sind wenig durchlässig. Man kann Undurchlässigkeit gelegentlich auch dort antreffen, wo zum Beispiel Frey (1990) eine aktive Schnittstelle zwischen Psychologie und experimenteller Ökonomie sieht: in der Diskussion um die Resultate des Heuristics-und-Biases-Programms. Weitgehend unbeachtet von der jeweils anderen Disziplin gibt es zur gleichen Zeit in der Psychologie (siehe Gigerenzer, 1996; Kahneman & Tversky, 1996) und in der Ökonomie eine Diskussion um die Validität der Befunde dieses Programms.

In der Ökonomie wird diese Diskussion unter den Auspizien einer Methodenkritik geführt, die mit guten Argumenten die experimentellen Routinen der Psychologie in Frage stellt. In der Psychologie häufen sich die Befunde, wonach sich die Prävalenz kognitiver Illusionen zum Teil enorm reduziert, wenn man die Repräsentation der Aufgaben ändert und ein frequentistisches anstelle eines probabilistischen Formats wählt (siehe Gigerenzer, 1997, für einen Überblick).¹¹ Diese Resultate der psychologischen Entscheidungsforschung scheinen in der experimentel-

len Ökonomie noch kaum Beachtung zu finden. Thaler (1991) zum Beispiel ist überzeugt, daß „mental illusions should be considered the rule rather than the exception“ (S. 4), und auch Rabin (in Druck), der für Ökonomen ein Resümee psychologischer Entscheidungsforschung zog, hat die kontroverse Diskussion in der Psychologie um die Existenz kognitiver Illusionen weitgehend ignoriert. Er schlußfolgerte, daß „researchers have documented many systematic departures from rationality in judgment under uncertainty“ (S. 23).

Resümiert man die Resultate zum frequentistischen Format (Psychologie) und zum Einfluß von Motivation und Lernen (experimentelle Ökonomie) auf menschliches Urteilen, dann kann man getrost Zweifel an Thalers (1991) Diktum, kognitive Illusionen seien die Regel, anmelden. Allerdings soll dieser Zweifel nicht so verstanden werden, daß der rationale Mensch der neoklassischen Ökonomie – ein Mensch, der, mit unbegrenztem Wissen und Verarbeitungskapazitäten ausgestattet, immerzu optimale Entscheidungen im Sinne der jeweiligen Nutzenfunktion trifft – damit empirisch plausibler würde. Die Herausforderung, die Herbert Simon (1955) in Gestalt der Unterscheidung zwischen substantiver (*substantive*) und prozeduraler (*procedural*, auch *bounded*) Rationalität an die ökonomische Theoriebildung gestellt hat, bleibt bestehen: Wo die Ökonomie Rationalität via Output (z.B. Wahlentscheidungen) der Erwartungsnutzentheorie konzeptualisiert, fordert Simon eine Theorie von Rationalität, die diese nicht auf den Output reduziert, sondern die Frage stellt, ob beides, Output *und* Prozeß, rational sind. Simon (1987) schreibt:

If ... we accept the proposition that both the knowledge and the computational power of the decision maker are severely limited, then we must distinguish between the real world and actor's perception of it and reasoning about it. That is to say, we must construct a theory (and test it empirically) of the processes of decision. ... The rational person of cognitive psychology goes about making his or her decisions in a way that is procedurally reasonable in the light of the available knowledge and means of computation. (S. 27)

Ogleich Thaler (1987a, S. 116; 1991, S. 4) wiederholt argumentierte, das Heuristics-und-Biases-Programm schreibt den Simonschen Begriff der prozeduralen Rationalität weiter, auch Camerer (1995, S. 588) die Wurzeln des Programms bei Simon sieht und gleichfalls Kahneman, Slovic und Tversky (1982) diese Beziehung herstellen, bleibt diese Ver-

¹¹ Theoretische Erklärungen für den reduzierenden Einfluß des Häufigkeitsformates auf das Overconfidence-Phänomen, die Mißachtung von Basisraten und den Konjunktionsfehler finden sich bei Gigerenzer et al. (1991), Gigerenzer und Hoffrage (1995) sowie Hertwig (1997).

wandtschaftsbeziehung fragwürdig (Gigerenzer, 1997). Das bedeutet aber auch, daß die Aufgabe, die Simon (1987) für die Ökonomie und Psychologie formuliert hat, prozedurale Rationalität besser zu verstehen, noch vor uns liegt.

Literatur

- Adair, J. G., Dushenko, T. W. & Lindsay, R. C. L. (1985). Ethical regulations and their impact on research practice. *American Psychologist*, *40*, 59–72.
- Akerlof, G. A. & Yellen, J. L. (1985). Can small deviations from rationality make significant differences to economic equilibria? *American Economic Review*, *75*, 708–720.
- Arkes, H. R., Dawes, R. M. & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, *37*, 93–110.
- Baumrind, D. (1964). Some thoughts on ethics of research. After reading Milgram's „Behavioral study of obedience.“ *American Psychologist*, *19*, 421–423.
- Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited. *American Psychologist*, *40*, 165–174.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, *96*, 85–94.
- Bonetti, S. (in press). Experimental economics and deception. *Journal of Economic Psychology*.
- Bower, B. (1997). Null science: Psychology's statistical status quo draws fire. *Science News*, *151*, 356–357.
- Campbell, D. J. (1984). The effects of goal-contingent payment on the performance of a complex task. *Personnel Psychology*, *37*, 23–40.
- Camerer, C. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *American Economic Review*, *77*, 981–997.
- Camerer, C. (1990). Do markets correct biases in probability judgment? Evidence from market experiments. In L. Green & J. Kagel (Eds.), *Advances in behavioral economics* (pp. 126–172). Greenwich, CT: JAI Press.
- Camerer, C. (1992). The rationality of prices and volume in experimental markets. *Organizational Behavior and Human Decision Processes*, *51*, 237–272.
- Camerer, C. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton University Press.
- Camerer, C. (1997). Rules for experimenting in psychology and economics, and why they differ. In W. Albers, W. Güth, P. Hammerstein, B. Moldovanu & E. van Damme (Eds.), *Understanding strategic interaction. Essays in honor of Reinhard Selten* (pp. 313–327). Heidelberg: Springer.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin*, *14*, 664–675.
- Cox, J. C. & Isaac, R. M. (1986). Experimental economics and experimental psychology: Ever the twain shall meet? In A. J. MacFadyen & H. W. MacFayden (Eds.), *Economic Psychology: Intersections in theory and application*. Amsterdam: North-Holland.
- Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.
- Davis, D. D. & Holt, C. A. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- Duh, R. R. & Sunder, S. (1986). Incentives, learning, and processing of information in a market environment: An examination of the base rate fallacy. In S. Moriarty (Ed.), *Laboratory market research*. Norman: University of Oklahoma Press.
- Edwards, W. (1961). Costs and payoffs are instructions. *Psychological Review*, *68*, 275–284.
- Eisenberger, R. & Cameron, J. (1996). Detrimental effects of reward. Reality or myth? *American Psychologist*, *51*, 1153–1166.
- Estes, W. K., Campbell, J. A., Hatsopoulos, N. & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 556–571.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, *50*, 123–129.
- Freeman, K. (1996, 6. Juni). Amos Tversky, expert on decision making, is dead at 59. *The New York Times Obituaries*, B16.
- Frey, B. S. (1990). Entscheidungsanomalien: Die Sicht der Ökonomie. *Psychologische Rundschau*, *41*, 67–83.
- Friedman, M. (1953). *Essays in positive economics*. Chicago: University of Chicago Press.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596.
- Gigerenzer, G. (1997). Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics*, *133*, 201–218.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.
- Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.
- Goodie, A. S. & Fantino, E. (1995). An experientially derived base-rate error in humans. *Psychological Science*, *6*, 101–106.

- Goodie, A. S. & Fantino, E. (1997). *Super-extended training and the arbitrary-identical effect*. Manuskript zur Veröffentlichung eingereicht.
- Grether, D. M. (1978). Recent psychological studies on behavior under uncertainty. *American Economic Review, Papers and Proceedings*, 68, 70–74.
- Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics*, 95, 537–557.
- Grether, D. M. (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization*, 17, 31–57.
- Grether, D. M. & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69, 623–638.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Gross, A. E. & Fleming, I. (1982). Twenty years of deception in social psychology. *Personality and Social Psychology Bulletin*, 8, 402–408.
- Harrison, G. W. (1994). Expected utility theory and the experimentalists. *Empirical Economics*, 19, 223–253.
- Hell, W., Gigerenzer, G., Gauggel, S., Mall, M. & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, 16, 533–538.
- Hertwig, R. (1997). Sind die Gesetze des Denkens die Gesetze der Wahrscheinlichkeitstheorie und Logik? In H. Mandl (Hrsg.), *Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie in München 1996*. Göttingen: Hogrefe.
- Hertwig, R. & Ortmann, A. (1997). *Experimental practices in economics: A methodological challenge for psychologists*. Manuskript zur Veröffentlichung eingereicht.
- Hey, J. D. (1991). *Experiments in economics*. Oxford: Basil Blackwell.
- Hey, J. D. (in press). Experimental economics and deception: A comment. *Journal of Economic Psychology*.
- Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, 90, 197–217.
- Hsee, C. K. & Weber, E. U. (1997). A fundamental prediction error: Self-others discrepancies in risk preference. *Journal of Experimental Psychology: General*, 126, 45–53.
- Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103, 582–591.
- Kliegl, R., Smith, J. & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, 25, 247–256.
- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, 19, 1–53.
- Krupat, E. & Garonzik, R. (1994). Subjects' expectations and the search for alternatives to deception in social psychology. *British Journal of Social Psychology*, 33, 211–222.
- Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 3–26.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.
- Lichtenstein, S., Fischhoff, B. & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Lindeman, S. T., van den Brink, W. P. & Hoogstraten, J. (1988). Effect of feedback on base-rate utilization. *Perceptual and Motor Skills*, 67, 343–350.
- Loewenstein, R. (1996, 13. Juni). Sure, markets are rational, just like life. *Wall Street Journal*, S. 1.
- Lopes, L. L. (1994). Psychology and economics: Perspectives on risk, cooperation, and the marketplace. *Annual Review of Psychology*, 45, 197–227.
- MacCoun, R. J. & Kerr, N. L. (1987). Suspicion in the psychological laboratory: Kelman's prophecy revisited. *American Psychologist*, 42, 199.
- Mellers, B. A. (1996). From the president. *J/DM Newsletter*, 15, 3.
- Mellers, B. A., Schwartz, A. & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology*, 49, 447–477.
- Neumann, J. von & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Newberry, B. H. (1973). Truth telling in subjects with information about experiments: Who is being deceived? *Journal of Personality and Social Psychology*, 25, 369–374.
- Nicks, S. D., Korn, J. H. & Mainieri, T. (1997). The rise and fall of deception in social psychology and personality research, 1921 to 1994. *Ethics & Behavior*, 7, 69–77.

- Ordóñez, L. D., Mellers, B. A., Chang, S. J. & Roberts, J. (1995). Are preference reversals reduced when made explicit? *Journal of Behavioral Decision Making*, 8, 265–277.
- Ortmann, A. & Hertwig, R. (1997). Is deception acceptable? *American Psychologist*, 52, 746–747.
- Phillips, L. D. & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72, 346–354.
- Rabin, M. (in press). Psychology and economics. *Journal of Economic Literature*.
- Robinson, J. (1977). What are the questions? *Journal of Economic Literature*, 15, 1318–1339.
- Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- Roth, A. E. (1995). Introduction to experimental economics. In J. Kagel & A. Roth (Eds.), *Handbook of experimental economics*, (pp. 3–109). Princeton, NJ: Princeton University Press.
- Roth, A. E. (1996). Individual rationality as a useful approximation: Comments on Tversky's 'Rational theory and constructive choice.' In K. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds.), *The rational foundations of economic behavior* (pp. 198–202). Macmillan.
- Russell, T. & Thaler, R. (1985). The relevance of quasi rationality in competitive markets. *American Economic Review*, 75, 1071–1082.
- Russell, R. R. & Wilkinson, M. (1979). *Microeconomics: A synthesis of modern and neoclassical theory*. New York: Wiley.
- Samuelson, P. (1997, 23. August). *The Economist*, S. 13.
- Sharpe, D., Adair, J. G. & Roesch, N. J. (1992). Twenty years of deception research: A decline in subjects' trust? *Personality and Social Psychology Bulletin*, 18, 585–590.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Simon, H. A. (1987). Rationality in psychology and economics. In R. M. Hogarth & M. W. Reder (Eds.), *Rational choice: The contrast between economics and psychology* (pp. 25–40). Chicago: The University of Chicago Press.
- Smith, C. P. & Berard, S. P. (1982). Why are human subjects less concerned about ethically problematic research than human subjects committees? *Journal of Applied Social Psychology*, 12, 209–221.
- Smith, S. S. & Richardson, D. (1983). Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology*, 44, 1075–1082.
- Smith, V. L. (1976). Experimental Economics: Induced value theory. *American Economic Review Proceedings*, 66, 274–279.
- Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72, 923–955.
- Smith, V. L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99, 877–897.
- Smith, V. L. & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31, 245–261.
- Taylor, K. M. & Shepperd, J. A. (1996). Probing suspicion among participants in deception research. *American Psychologists*, 51, 886–887.
- Thaler, R. (1987a). The psychology of choice and the assumptions of economics. In A. E. Roth (Eds.), *Laboratory experimentation in economics: Six points of view* (pp. 99–130). Cambridge: Cambridge University Press.
- Thaler, R. H. (1987b). The psychology and economics conference handbook: Comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman. In R. M. Hogarth & M. W. Reder (Eds.), *Rational choice: The contrast between economics and psychology* (pp. 95–100). Chicago: The University of Chicago Press.
- Thaler, R. H. (1991). *Quasi rational economics*. New York: Russel Sage Foundation.
- Thurstone, L. L. (1931). The indifference function. *Journal of Social Psychology*, 2, 139–167.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- Tversky, A. & Kahneman, D. (1987). Rational choice and the framing of decisions. In R. M. Hogarth & M. W. Reder (Eds.), *Rational choice: The contrast between economics and psychology* (pp. 67–94). Chicago: The University of Chicago Press.
- Wallis, W. A. & Friedman, M. (1942). The empirical derivation of indifference functions. In O. Lange, F. McIntyre & T. O. Yntema (Eds.), *Studies in mathematical economics and econometrics in memory of Henry Schultz*. Chicago: University of Chicago Press.
- Winkler, R. L. & Murphy, A. H. (1973). Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, 10, 252–270.
- Winterfeldt, D. von & Edwards, W. (1982). Costs and payoffs in perceptual research. *Psychological Bulletin*, 91, 609–622.
- Wright, W. F. & Aboul-Ezz, M. E. (1988). Effects of extrinsic incentives on the quality of frequency assessments. *Organizational Behavior and Human Decision Processes*, 41, 143–152.
- Wright, W. F. & Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the an-

choring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes*, 44, 68–82.

Wundt, W. (1863). *Vorlesungen über die Menschen- und Thierseele*, Bd. 1. Leipzig: Leopold Voß.

Dr. Ralph Hertwig

Max-Planck-Institut für Bildungsforschung
Lentzeallee 94
D-14195 Berlin
E-mail: hertwig@mpib-berlin.mpg.de

