# Die Newcomb-Paradoxie – und ihre Lösung*

von Wolfgang Lenzen

*Universität Osnabrück*

## 1 Einleitung

In der Festschrift zum 65. Geburtstag von Carl G. Hempel veröffentlichte Robert Nozick einen vielbeachteten Aufsatz, in dem er die philosophische Leserschaft mit einem Entscheidungsproblem bekannt machte, das ursprünglich vom kalifornischen Physiker William Newcomb erfunden worden war. Mir ist nicht bekannt, was Hempel seinerzeit von Nozicks Artikel gehalten hat. Ich hoffe aber, daß Franz von Kutschera meinen Beitrag zur Feier seines 65. Geburtstag als korrekte Lösung der Newcomb-Paradoxie akzeptiert.[1] In den vergangenen knapp 30 Jahren hat die Arbeit von Nozick eine kaum überschaubare Lawine von Aufsätzen ausgelöst, und die Anzahl der Lösungsvorschläge ist Legion. Das Newcomb-*Problem* besteht darin, zu entscheiden, welche von zwei Handlungsweisen in der gleich zu schildernden Situation die rationale oder richtige ist. Auf diese Frage gibt es trivialerweise nur zwei Antworten. Die Newcomb-*Paradoxie* hingegen besteht darin, daß für beide Entscheidungen prima facie sehr überzeugende Argumente existieren. Eine *Lösung* muß dann entweder zeigen, daß die Prämissen der Newcomb-Situation in irgendeinem Sinne inkonsistent sind; oder es muß nachgewiesen werden, daß die Argumentation zugunsten einer Handlung eben nur scheinbar überzeugend, in Wirklichkeit aber unschlüssig ist.

## 2 Das Problem

Die von Nozick [1969: 114/5] formulierte Version des Entscheidungsproblems lautet (in meiner Übersetzung):

> „Stell Dir ein Wesen vor, in dessen Fähigkeit, Deine Entscheidungen vorherzusagen, Du größtes Vertrauen besitzt. (Man könnte eine Science-Fiction-Story von einem Wesen aus einer anderen Welt mit fortgeschrittener Technik und Wissenschaft erfinden, das, wie Du weißt, Dir freundlich gesonnen ist, usw.). Du weißt, daß dieses We-

* Dank an Christoph Lumer und die Teilnehmer unseres gemeinsamen Newcomb-Seminars im WS 1995/96 für lebhafte und fruchtbare Diskussionen; Dank auch an Stephan Guhe für wertvolle kritische Anregungen!

[1] Kutschera selber hat sich zwar mit anderen Paradoxien im Umfeld der *Logik der Normen, Werte und Entscheidungen* beschäftigt, nicht aber mit dem Newcomb-Paradox, das zur Zeit der Abfassung seines Buchs [1973] noch nicht hinreichend rezipiert worden war. Die erste Auseinandersetzung im deutschsprachigen Raum findet sich anscheinend bei Spohn [1978].

sen Deine Entscheidungen in der Vergangenheit oft korrekt vorhergesagt hat (und, so weit Du weißt, hat es Deine Entscheidungen niemals falsch vorhergesagt); außerdem weißt Du, daß das Wesen die Entscheidungen von anderen Leuten, die Dir sehr ähnlich waren und die sich in der speziellen, gleich zu schildernden Situation befunden haben, oft korrekt vorhergesagt hat. Man könnte die Geschichte noch auswalzen, jedenfalls bringt all dies Dich dazu zu glauben, daß die Vorhersage des Wesens über Deine Entscheidung in der zu diskutierenden Situation mit annähernder Sicherheit korrekt sein wird.
Es gibt zwei Kisten, $K_1$ und $K_2$. $K_1$ enthält 1.000 \$, $K_2$ enthält entweder 1.000.000 \$ (M \$) oder nichts. Wovon es abhängt, welcher Betrag sich in $K_2$ befindet, wird gleich gesagt. [...] Du hast die Wahl zwischen zwei Handlungen:
(1) Nimm den Inhalt beider Kisten
(2) Nimm nur, was sich in der zweiten Kiste befindet.
Darüber hinaus gilt, wie Du weißt, und wie das Wesen weiß, daß Du es weißt, etc.:
(I) Wenn das Wesen vorhersagt, daß Du den Inhalt beider Kisten nimmst, dann legt es die M \$ nicht in die zweite Kiste
(II) Wenn das Wesen vorhersagt, daß Du nur den Inhalt der zweiten Kiste nimmst, dann legt es die M \$ in die zweite Kiste.[2]
Die Vorgehensweise ist wie folgt: Zunächst macht das Wesen seine Vorhersage. Dann legt es, oder legt nicht, die M \$ in die zweite Kiste, je nachdem, was es vorhergesagt hat. Danach triffst Du Deine Entscheidung. Was sollst Du tun?“

Was jemand vernünftigerweise *tun* sollte, hängt entscheidend davon ab, was er *will* und was er *glaubt*. Die Frage des Wollens dürfte dabei wenig kontrovers sein. Es ist plausibel anzunehmen, daß die Präferenzen der Handelnden eindeutig durch die Höhe des zu gewinnenden Geldbetrags bestimmt werden (wenngleich in manchen Diskussionen des Problems auch in Anschlag gebracht wurde, daß jemand mehr als am schnöden Mammon daran interessiert sein könnte, seine menschliche Freiheit zu beweisen)[3]. Alles andere als unkontrovers ist hingegen die Frage, was man in der Newcomb-Situation vernünftigerweise *glauben kann* bzw. *glauben sollte*. Die im folgenden präsentierten Argumente scheinen nämlich überzeugend zu begründen, daß es einerseits besser ist, nur die Kiste $K_2$ zu wählen, während es andererseits auch besser erscheint, beide Kisten zu nehmen. So etwas kann jedoch kein rationales Subjekt zugleich glauben.

## 3 Die Argumente

[2] Nozick ergänzt hier per Fußnote: „Wenn das Wesen vorhersagt, daß Du Deine Entscheidung von einem Zufallsexperiment abhängen läßt, z.B. durch Wurf einer Münze [...], dann legt es die M \$ nicht in die zweite Kiste.“ Auf diese Komplikation gehe ich im folgenden nicht ein, sondern setze der Einfachheit halber voraus, daß sich jeder Spieler eindeutig für die eine oder für die andere Handlung entscheidet.

[3] Vgl. die Ausführungen von I. Asimov, die in Dacey et al. [1977: 80] zitiert werden.

Das Argument zugunsten der Entscheidung, nur die Kiste $K_2$ zu nehmen, stützt sich auf die vorausgesetzte hohe Zuverlässigkeit der Vorhersage:

> „Wenn ich den Inhalt beider Kisten nehme, wird das Wesen dies mit annähernder Sicherheit vorhergesagt und die M $ nicht in die zweite Kiste getan haben; deshalb werde ich mit annähernder Sicherheit nur 1.000 $ bekommen. Wenn ich nur das nehme, was sich in der zweiten Kiste befindet, wird das Wesen dies mit annähernder Sicherheit vorhergesagt und die M $ in die zweite Kiste getan haben; deshalb werde ich mit annähernder Sicherheit 1.000.000 $ bekommen. [...] Deshalb sollte ich nur das nehmen, was sich in der zweiten Kiste befindet." (Nozick 1969: 115)

Das konkurrierende Argument beruft sich statt dessen auf die Tatsache, daß zum Zeitpunkt der Entscheidung bereits feststeht, wieviel Geld in $K_2$ gelegt wurde und daß die Entscheidung des Handelnden hieran nichts mehr zu ändern vermag:

> „Wenn das Wesen die M $ bereits in die zweite Kiste gelegt hat, dann erhalte ich M $ plus 1.000 $, sofern ich den Inhalt beider Kisten nehme, wohingegen ich nur M $ bekomme, wenn ich nur das nehme, was in der zweiten Kiste ist. Wenn das Wesen die M $ nicht in die zweite Kiste gelegt hat und wenn ich den Inhalt beider Kisten nehme, dann bekomme ich 1.000 $, wohingegen ich überhaupt kein Geld bekomme, wenn ich nur das nehme, was in der zweiten Kiste ist. Egal ob die M $ drin sind oder nicht, was bereits fest determiniert ist, bekomme ich also, wenn ich den Inhalt beider Kisten nehme, 1.000 $ mehr als wenn ich nur das nehme, was in der zweiten Kiste ist. Deshalb sollte ich nehmen, was in beiden Kisten ist." (ibid.)

Zur Vereinfachung der folgenden Diskussion ein paar Abkürzungen:

> $S_1$ := Zum Zeitpunkt der Entscheidung liegen 1.000.000 $ in $K_2$;
> $S_2$ := Zum Zeitpunkt der Entscheidung liegen 0 $ in $K_2$;
> $H_1$ := Der Handelnde nimmt nur den Inhalt von $K_2$;
> $H_2$ := Der Handelnde nimmt den Inhalt beider Kisten.

Der Kern des Newcomb-Problems läßt sich dann folgendermaßen zusammenfassen: Es gibt zwei mögliche Situationen $S_1$ und $S_2$. In beiden Fällen führt die Handlung $H_2$ zu einem besseren Resultat als $H_1$, d.h. – in der Terminologie der Entscheidungstheorie – $H_2$ ist gegenüber $H_1$ stark *dominant*. Trotzdem erscheint es insgesamt besser, die nicht-dominante Handlung $H_1$ zu tun, weil hierdurch anscheinend sichergestellt wird, daß die Situation $S_1$ eintritt (was zugleich den Gewinn von 1.000.000 $ bedeutet), während die dominante Handlung $H_2$ scheinbar dazu führt, daß die Situation $S_2$ vorliegt und somit nur einen Gewinn von 1.000 $ verspricht.

**4 Entscheidungstheorie**

Die vorangehende Beschreibung des Dilemmas legt es nahe, das Newcomb-Problem als einen Konflikt zwischen zwei allgemeinen entscheidungstheoretischen Prinzipien auffassen, nämlich zwischen dem Dominanzprinzip:

(DOM) Wenn in einer Entscheidungssituation eine (schwach oder stark) dominante Handlungsalternative H existiert, dann ist es auf jeden Fall rational, H zu tun

auf der einen und dem Prinzip der Maximierung des zu erwartenden Nutzens auf der anderen Seite:

(MAX) Wähle stets jene Handlung, deren zu erwartender Nutzen (im Vergleich zum Nutzen der alternativen Handlungen) maximal ist.

Dieser Konflikt an sich stellt jedoch keine Paradoxie dar. Obgleich man auf den ersten Blick erwarten würde, daß eine dominante Handlung *immer* zugleich den größten zu erwartenden Nutzen hat und also mit (MAX) in Einklang steht, ist dies nicht allgemein der Fall. Wie das folgende Beispiel aus Nozick [1969: 123] verdeutlicht, kann es insbesondere dann rational sein, eine nicht-dominante Handlung $H_1$* zu wählen, wenn es *von den jeweiligen Handlungen abhängt*, welche der fraglichen Situationen $S_1$* und $S_2$* eintritt:

> „Angenommen, ich spiele an einem manipulierten Roulettetisch und der Besitzer des Casinos bietet mir die Wahl zwischen den Handlungen $H_1$* und $H_2$* an, für die die folgende Auszahlungsmatrix gilt (wobei $S_1$* bedeutet ‘Beim nächsten Durchgang kommt Schwarz’ und $S_2$* = ‘Beim nächsten Durchgang kommt Rot’):
>
> | | $S_1$* | $S_2$* |
> |---|---|---|
> | $H_1$* | 5 $ | 90 $ |
> | $H_2$* | 10 $ | 100 $ |
>
> Angenommen auch noch, ich wüßte, daß der Croupier, der das Roulette bedient und der [...] sich gegenüber dem Besitzer vollkommen loyal verhält, angewiesen wurde, beim nächsten Durchgang Schwarz kommen zu lassen, wenn ich mich für $H_2$* entscheide, aber Rot kommen zu lassen, falls ich $H_1$* wähle. Obwohl $H_2$* gegenüber $H_1$* die dominante Handlung ist, sollte ich angesichts meiner Kenntnis dieser Situation offensichtlich $H_1$* wählen.“

Der Grund für die Rationalität dieser Entscheidung liegt darin, daß die Situationen $S_1$* und $S_2$* von den Handlungen $H_1$* und $H_2$* *kausal beeinflußt* werden bzw. – wie Nozick [1969: 123] es vorsichtiger formuliert – von ihnen „nicht wahrscheinlichkeitstheoretisch unabhängig sind". Auf der abstrakten Ebene allgemeiner entscheidungstheoretischer Prinzipien steht das zur Stützung von $H_1$ herangezogene (MAX) also besser da als das für $H_2$ benutzte (DOM), welches auf Entscheidungssituationen eingeschränkt werden müßte, bei denen die Handlungen keinerlei kausalen Einfluß darauf haben, welcher der für die Auszahlungen relevanten Zustände eintritt. Wie solche Einschränkungen im Einzelnen ausschauen müßten, braucht hier jedoch nicht näher diskutiert zu werden[4]. Denn die Plausibilität des obigen Argumentes zugunsten von $H_2$ beruht ja gerade darauf, daß die jeweiligen Handlungen in der Newcomb-Situation keinerlei Einfluß mehr darauf haben, ob $S_1$ oder $S_2$ vorliegt. Im folgenden soll nun gezeigt werden, daß – bzw. in welchem Sinne – die dominante Handlung $H_2$ auch angesichts des konkurrierenden Arguments für $H_1$ als *die* „rationale" Entscheidung verteidigt werden kann.

[4] Ebenso brauchen wir hier nicht auf die im Anschluß an das Newcomb-Problem entwickelten „kausalen Entscheidungstheorien" z.B. von Gibbard/Harper [1978] und Lewis [1981] einzugehen.

## 5 Rationale Entscheidung und rationaler Glaube

Im üblichen Verständnis der Entscheidungstheorie hängt die Rationalität einer Handlung nur von den Wünschen und dem *Glauben* der jeweiligen Person x ab, wie vernünftig oder irrational dieser Glaube selber auch immer sein mag[5]. In einem strengeren Sinn würde man jedoch nur solche Handlungen als rational bezeichnen wollen, bei denen die für die Entscheidung maßgebenden Glaubensannahmen hinreichend begründet sind. Warum bzw. mit welchem Grund sollte man aber an die konkurrierenden Prämissen des Newcomb-Problems glauben? Zur Begründung der einen Prämisse darf man voraussetzen, daß ein neutraler „Schiedsrichter" über die Einhaltung der Spielregeln achtet und kontrolliert, daß in jedem Einzelfall *zunächst* der Vorhersager seine Vorhersage trifft und das Geld in $K_2$ deponiert bzw. nicht deponiert, und daß erst *anschließend* die jeweilige Person ihre Entscheidung trifft, ohne daß hierdurch das Bestehen oder Nicht-Bestehen von $S_1$ bzw. $S_2$ noch verändert werden könnte. Was die andere Prämisse betrifft, so wurde im ursprünglichen Szenario angenommen, daß man aufgrund recht obskurer Hinweise von der Unfehlbarkeit des Vorhersagers überzeugt sein möge. Nun ist aber ein Glaube an Science-Fiction-Stories und Wesen von einem anderen Stern nicht jedermanns Sache. Deshalb sollte man lieber zur folgenden, von Nozick [1969: 115/116] formulierten Variante übergehen, bei der die Zuverlässigkeit des Vorhersagers auf empirischen Informationen beruht:

> „Du weißt, daß viele Leute Deines Schlags, Philosophie-Studenten, Professoren, usw. das Experiment mitgemacht haben. All diejenigen, die nur den Inhalt der zweiten Kiste nahmen, auch solche, die das zweite Argument [für $H_2$] zwar kannten, aber nicht befolgten, erhielten 1.000.000 $. Ferner weißt Du, daß all die cleveren Burschen, die dem zweiten Argument folgten und den Inhalt beider Kisten nahmen, nur 1.000 $ bekamen."

Bei der ursprünglichen Fassung war jedoch nur davon die Rede gewesen, daß das fragliche Wesen die Entscheidungen „*oft*" korrekt vorhergesagt habe bzw. daß seine Vorhersagen „mit *annähernder* Sicherheit" korrekt wären. Die gerade zitierte Spielart sieht stattdessen vor, daß die Vorhersagen des Wesens *immer* korrekt sind. Beide Varianten müssen jedoch separat behandelt werden. Der Fall der absolut sicheren Vorhersage wird erst in Abschnitt 8 behandelt.

Für den jetzt zu diskutierenden Fall der annähernd sicheren Vorhersage sei der Einfachheit halber angenommen, es hätten bislang 200 Personen am Spiel teilgenommen, darunter 100 „Gläubige" $G = \{x_1,...,x_{100}\}$, die sich alle für $H_1$ entschieden, und 100 „Skeptiker" S =

[5] Vgl. z.B. Eells [1985: 198]: „An act is rational relative to a possessed body of information; the quality of the information and the actual facts are both irrelevant."

$\{y_1,...,y_{100}\}$, die alle $H_2$ wählten. Ferner mögen 99 % der $x_i$ in den Genuß von 1.000.000 $ gekommen sein, während einer – sagen wir x* – Pech hatte und leer ausging. Entsprechend mögen 99 % der $y_j$ lediglich 1.000 $ erhalten haben, während wiederum einer, y*, das ganz große Los zog und mit 1.001.000 $ von dannen ziehen konnte. Bei Kenntnis dieser Daten (und in Abwesenheit gegenteiliger Informationen) sollte jeder rational Glaubende z diese relativen Häufigkeiten bzw. „objektiven Wahrscheinlichkeiten" als seine eigenen, subjektiven Wahrscheinlichkeiten übernehmen. Wenn z also erfährt, daß sich die Person $x_i$ bei dem gerade stattfindenden Newcomb-Experiment für $H_1$ entschieden hat, so wäre es für z rational, mit hoher Sicherheit zu glauben bzw. darauf zu wetten, daß bei $x_i$ die Situation $S_1$ vorliegt. Umgekehrt sollte z, wenn er erfährt, daß bei einem Durchgang die Situation $S_1$ vorliegt, mit hoher Sicherheit erwarten, daß $x_i$ sich für $H_1$ entscheiden wird. Und analog natürlich hinsichtlich $H_2$ und $S_2$. Insgesamt erscheint es somit für z vollkommen rational, in hohem Grad davon überzeugt zu sein, daß eine beliebige Person sich genau dann für $H_1$ entscheidet, wenn die Situation $S_1$ vorliegt, und genau dann für $H_2$, wenn $S_2$. Was für *jede* rationale Person vernünftig zu glauben ist, das sollte aber vernünftigerweise auch ein *Skeptiker* glauben. Wie konnte dann aber seine Entscheidung für $H_2$ jemals *rational* gewesen sein? Im folgenden Abschnitt wird zunächst gezeigt, daß Skeptiker zumindest in dem Sinne „rational" handeln, als $H_2$ für alle Akteure die *in der jeweiligen Entscheidungssituation gewinnmaximierende* Handlungsalternative darstellt. Dazu betrachten wir eine nützliche Fiktion:

## 6 Der Beobachter

Gemäß Nozick [1969: 116] darf man das Newcomb-Szenario ohne Verfälschung der relevanten Konstruktionsmerkmale wie folgt modifizieren:

> „Das Wesen hat seine Vorhersage bereits getroffen und die M $ in der zweiten Kiste deponiert oder nicht deponiert [...] Dies geschah vor einer Woche; [...] Die M $ liegen bereits in der Kiste $K_2$ bzw. nicht [...] [O]bwohl Du von Deiner Seite aus nicht in $K_2$ schauen kannst, so ist sie doch auf der anderen Seite durchsichtig. Auf dieser anderen Seite von $K_2$ sitze ich, schaue hinein und sehe, was drin ist. Entweder schaue ich bereits seit einer Woche auf die M $, oder ich schaue bereits seit einer Woche auf eine leere Kiste. Wenn das Geld schon drin ist, bleibt es drin, wie immer Du Dich entscheidest. Es wird nicht verschwinden. Wenn es nicht bereits drin ist, wenn ich also auf eine leere Kiste schaue, dann wird es nicht plötzlich auftauchen, falls Du Dich entscheidest, nur den Inhalt der zweiten Kiste zu nehmen.

Ein solcher Beobachter, der quasi als Schiedsrichter über die Einhaltung der Spielregeln wacht, möge auch schon bei den Spielern $\{x_1,...,x_{100}\}$ und $\{y_1,...,y_{100}\}$ präsent gewesen sein. Dann hätte er das objektive Geschehen wie folgt beschrieben: In unregelmäßiger Folge lag insgesamt 100-mal die Situation $S_1$ und ebenso oft die Situation $S_2$ vor, d.h. 100-mal befanden sich in der Kiste $K_2$ 1.000.000 \$ und 100-mal war $K_2$ leer. In 99 % der Fälle von $S_1$ betrat ein Gläubiger das Versuchszimmer, überlegte eine Weile, ließ die 1.000 \$ in $K_1$ unberührt und nahm nur den Inhalt von $K_2$. Entsprechend kam in 99 % der Fälle von $S_2$ ein Skeptiker in den Raum, überlegte eine Weile und nahm (außer der leeren Kiste $K_2$) die 1.000 \$ aus $K_1$. Nur ein einziges Mal tauchte in der Situation $S_1$ ein Skeptiker, sagen wir y*, auf, der daraufhin 1.0001.000 \$ kassierte; und nur ein einziges Mal tauchte in der Situation $S_2$ ein Gläubiger, x*, auf, der ohne irgendeinen \$ von dannen ziehen mußte. Diese *objektive Beschreibung der Fakten* rechtfertigt nicht nur die beiden singulären irrealen Konditionalsätze

(IK1) Wenn x* beide Kisten genommen hätte, so hätte er zumindest 1.000 \$ erhalten

(IK2) Wenn y* bloß die Kiste $K_2$ genommen hätte, so hätte er „nur" 1.000.000 \$ bekommen

sondern auch die folgenden generalisierten Aussagen für beliebige $x_i \in G$ und $y_j \in S$ mit $x_i \neq x^*$ und $y_j \neq y^*$:

(IK3) Wenn $x_i$ beide Kisten genommen hätte, dann hätte er 1.001.000 \$ bekommen

(IK4) Wenn $y_j$ nur die Kiste $K_2$ gewählt hätte, dann hätte er gar nichts bekommen.

Trotz der wichtigen Untersuchungen von Stalnaker [1968], Lewis [1973] und Kutschera [1974; 1976] sind die Wahrheitsbedingungen für Konditionalsätze leider immer noch recht vage. Relativ unkontrovers ist lediglich die Einsicht, daß die Wahrheit oder Akzeptabilität eines irrealen Konditionalsatzes typischerweise auf entsprechende *Gesetzesaussagen* rekurriert. Im vorliegenden Fall handelt es sich dabei nicht um ein Naturgesetz, sondern um die analytisch geltende „Spielregel", die man mit Hilfe der Parameter

$t_G$ = der Zeitpunkt der Gelddeponierung durch den Vorhersager
$t_E$ = der Zeitpunkt der Entscheidung des Handelnden
$t_A$ = der Zeitpunkt der Auszahlung des Gewinns

für beliebige Akteure z wie folgt präzisieren kann:

(SR) (1) Wenn zum Zeitpunkt $t_G$ die Situation $S_1$ vorliegt, dann gilt:
{(wenn z sich zu $t_E$ für $H_1$ entscheidet, so erhält er zu $t_A$ 1.000.000 \$) und (wenn z sich zu $t_E$ für $H_2$ entscheidet, so erhält er 1.001.000 \$)} und
(2) Wenn zum Zeitpunkt $t_G$ die Situation $S_2$ vorliegt, dann gilt:
{(wenn z sich zu $t_E$ für $H_1$ entscheidet, dann erhält er zu $t_A$ 0 \$) und (wenn z sich zu $t_E$ für $H_2$ entscheidet, so erhält er 1.000 \$)}.

Ebenso wie aus (SR) zusammen mit den *Fakten*, welche Situation bei welcher Person vorgelegen und für welche Handlung sie sich entschieden hat, logisch folgt, wer *de facto* wieviel Geld gewonnen hat, so folgt aus (SR) völlig analog, wieviel Geld welche Person *kontrafak-*

*tisch* gewonnen hätte, wenn sie sich – *in der gleichen Situation* – anders entschieden hätte. Eben diese logischen Folgerungen bilden die Basis für die Wahrheit bzw. Akzeptierbarkeit von (IK1) – (IK4) und damit auch der Schlußfolgerung, daß nicht nur y*, sondern ebenso gut jeder andere Skeptiker $y_j$ in der Situation, in der er sich befand, *gewinnmaximierend* gehandelt hat. Hingegen war die Entscheidung nicht nur des Pechvogels x*, sondern auch der anderen Gläubigen $x_i$ in der Situation, in der sie sich befanden, *suboptimal* (wiewohl der Verlust von zusätzlichen 1.000 $ angesichts der gewonnenen 1.000.000 $ praktisch nicht ins Gewicht fiel).

## 7 Eine konditionalsatzlogische Paradoxie?

Gegen diesen Lösungsansatz haben Verfechter der $H_1$-Strategie[6] eingewandt, daß die andere Prämisse der Newcomb-Situation nicht adäquat berücksichtigt wird. Ob ein Handelnder z sich in der Situation $S_1$ oder in $S_2$ befindet, sei doch keine Frage von Zufall und Glück, sondern hänge gerade davon ab, ob das mächtige Wesen vorhergesagt hat, daß z sich mit $K_2$ zufriedengeben wird. Aufgrund dieses „Gesetzes“ sollte man vielmehr die folgenden irrealen Konditionalsätze (wiederum für beliebige $x_i \in G$ und $y_j \in S$ mit $x_i \neq x^*$ und $y_j \neq y^*$) für wahr ansehen:

(IK5) Wenn $x_i$ beide Kisten gewählt hätte, dann hätte W dies vorhergesagt und demzufolge keine 1.000.000 in $K_2$ gelegt;

(IK6) Wenn $y_j$ nur die Kiste $K_2$ gewählt hätte, dann hätte W dies vorhergesagt und demzufolge 1.000.000 in $K_2$ gelegt.

Zusammen mit den übrigen Spielregeln ergäbe sich dann:

(IK7) Wenn $x_i$ beide Kisten gewählt hätte, dann hätte er nur 1.000 $ bekommen;

(IK8) Wenn $y_j$ nur die Kiste $K_2$ gewählt hätte, dann hätte er 1.000.000 $ bekommen.

Auch wenn – wie oben betont wurde – immer noch kein allgemein akzeptiertes Standardsystem der Konditionalsatzlogik existiert, so dürfte außer Frage stehen, daß die beiden letzten Sätze (IK7) und (IK8) mit den früheren (IK3) und (IK4) absolut *unverträglich* sind. Wenn sich also (IK5) – (IK8) ähnlich zwingend begründen ließen, wie (IK1) – (IK4) weiter oben durch die Spielregel (SR) begründet wurden, dann läge insgesamt eine zwar nicht strikt *logische*, aber immerhin *konditionalsatzlogische* Paradoxie vor. Denn dann wären im Newcomb-Szenario sowohl die Aussagen (IK3) und (IK4) als auch (IK7) und (IK8) zusammen wahr. Wie ist es aber um die Begründung von (IK5) bzw. (IK6) – und also ihrer Korollare (IK7) und (IK8) – bestellt?

[6] Vgl. z.B. Bar-Hillel/Margalit [1972], Olin [1976] und Horgan [1981].

Unter Bezugnahme auf die von Lewis [1973] entwickelte Semantik für irreale Konditionalsätze hat Horgan [1981: 333 ff.] zu zeigen versucht, daß z.B. (IK7) ebenso gut begründet sei wie (IK3). Irreale Konditionalsätze seien grundsätzlich vage, wobei sich die Vagheit auf zwei verschiedene Arten auflösen lasse. Man betrachte z.B. die Situation eines Gläubigen $x_i$, der sich de facto in der Situation $S_1$ befand und erwartungsgemäß $H_1$ wählte. Bei der „Standard“-Auflösung soll der Sachverhalt, daß der Vorhersager 1.000.000 $ in $K_2$ gelegt hat, auch in der „der wirklichen Welt ähnlichsten“ Welt, die für die Ermittlung des Wahrheitswertes einer irrealen Konditionaussage à la Lewis in Betracht zu ziehen ist, bestehen bleiben. Wenn also im Antezedens von (IK3) kontrafaktisch angenommen wird, daß $x_i$ sich für $H_2$ entschieden hätte, würde bei „Standard“-Auflösung weiterhin $S_1$ gelten, so daß – wie im Sukzedens von (IK3) behauptet wird – $x_i$ 1.001.000 $ bekommen hätte. Daneben existiert aber auch die Möglichkeit einer „Nicht-Standard“-Auflösung, bei der die Vorhersage-Zuverlässigkeit des Wesens einen wichtigeren Parameter für die Ähnlichkeit zwischen Welten darstellen soll als die Übereinstimmung in den kontingenten Fakten des Einzelfalls.[7] In diesem Fall wäre die ähnlichste Welt, in der $x_i$ beide Kisten nimmt, eine solche, in der das Wesen dieses korrekt vorhergesagt hat und in der somit das Sukzedens von (IK5) bzw. (IK7) wahr wird.

Horgan übersieht aber, daß der in Lewis [1979: 472] entwickelte Grundgedanke für die Bestimmung der „Ähnlichkeit“ zwischen Welten – „It is of the first importance to avoid big, complicated, varied, widespread violations of law“ – wirkliche (Natur-)*Gesetze* im Auge hat und nicht bloß statistische Korrelationen. Die Akzeptabilität oder Wahrheit von (IK5), (IK6) steht und fällt mit der Annahme der *Unfehlbarkeit* des mächtigen Wesens W. Bei der bislang diskutierten Variante einer mit 99 % zwar sehr hohen, aber eben nicht absolut sicheren Zuverlässigkeit besteht keine Möglichkeit zu folgern, daß wenn irgendjemand sich anders entschieden hätte als er sich de facto entschieden hat, dann das mächtige Wesen dies vorhergesehen hätte: Der Fall des Pechvogels x* bzw. des Maximalgewinners y* zeigt dies aufs deutlichste. Ganz grundsätzlich bleibt festzustellen, daß eine noch so hohe probabilistische Korrelation zwischen Ereignisarten A und B einen entsprechenden irrealen Konditionalsatz niemals zu stützen vermag. Daß auf A mit großer Wahrscheinlichkeit B folgt, gestattet zwar den Schluß, daß dem Ereignis Non-B mit großer Wahrscheinlichkeit Non-A vorausgegangen ist. Aber wenn in einem bestimmten Einzelfall von A erwartungsgemäß B eingetreten ist, darf man

[7] Vgl. Horgan [1981: 337]: „... the being’s predictive correctness is a more important parameter of similarity than is maximization of the spatiotemporal region through which perfect match of particular fact prevails“.

keineswegs schließen, daß wenn B nicht eingetreten wäre, dann auch A nicht stattgefunden hätte. Schließlich und endlich ist das unwahrscheinliche Ereignis (A und Non-B) immer *möglich*! Man betrachte z.B. den Wurf von sechs fairen Würfeln! Die Wahrscheinlichkeit dafür, daß alle sechs Würfel die gleiche Zahl zeigen, beträgt $(1/6)^5$, also ca. 0,00013. Aber wenn während eines langen Würfelabends doch einmal ein solch seltener Wurf gelingen sollte, wäre es absurd, kontrafaktisch schließen zu wollen, daß es sich nicht um den Wurf von sechs fairen Würfeln gehandelt habe. Um – in Anwendung auf das Newcomb-Problem – die fraglichen Konditionalsätze (IK5), (IK6) zu begründen, müßte der Verfechter der $H_1$-Strategie also zur zweiten Variante übergehen, d.h. dem mächtigen Wesen die Fähigkeit einer *absolut sicheren* Vorhersage unterstellen.

## 8 Der unfehlbare Vorhersager

Im folgenden sei also vorausgesetzt, daß für beliebige Spieler z gilt:

(U) (z entscheidet sich für $H_1$ gdw. W vorhergesagt hat, daß z $H_1$ wählt) und
(z entscheidet sich für $H_2$ gdw. W vorhergesagt hat, daß z $H_2$ wählt).

Nun ist aber seit dem *Neuen Rätsel der Induktion* von Goodman [1955] klar, daß selbst eine strikte Allaussage der Form $\Lambda z(Fz \supset Gz)$ (bzw. $\Lambda z(Fz \equiv Gz)$) den irrealen Konditionalsatz ‘Wenn z kein G wäre, dann wäre z auch kein F’ nur dann zu stützen vermag, wenn sie *gesetzartigen* Charakter hat. Deshalb muß der Proponent von (IK5), (IK6) weiterhin plausibel machen, daß es sich bei (U) um eine gesetzartige und nicht bloß akzidentell wahre Allaussage handelt. Dies ließe sich nach Nozick [1969: 134] durch folgende Zusatzannahme sicherstellen:

> „Er [der Vorhersager] beobachtet Dich einige Zeit bevor Du die Entscheidung triffst; er untersucht Dich mit einem komplizierten Apparat usw., und anschließend benutzt er seine Theorie, um auf der Grundlage Deines früheren Zustandes vorherzusagen, welche Wahl Du in der späteren Entscheidungssituation treffen wirst.“

Zusätzlich zu den früher eingeführten Parametern sei

$t_U$ = der Zeitpunkt der Untersuchung durch den Vorhersager

wobei natürlich $t_U < t_G < t_E < t_A$. Damit nun die Theorie des Vorhersagers in der Form (U) funktionieren kann, müßten die für die Prognose herangezogenen Zustände der Spieler zu $t_U$ auch mit den Entscheidungen bzw. Handlungen zu $t_E$ gesetzartig verknüpft sein. Dies könnte man durch die zusätzliche Annahme des folgenden „Gesetzes“ sicherzustellen versuchen:

(D) Für alle z gilt: {(z entscheidet sich zu $t_E$ für $H_1$ gdw. z sich zu $t_U$ im Zustand $Z_1$ befindet) und (z entscheidet sich zu $t_E$ für $H_2$ gdw. z sich zu $t_U$ im Zustand $Z_2$ befindet)}.[8]

Läßt man einmal außer Acht, welche Gründe jemals für die Wahrheit von (D) und von (U) sprechen könnten[9], so muß man feststellen, daß bei Annahme eines starken Determinismus à la (D) das Newcomb-Problem als Entscheidungsproblem *jeglichen Sinn verliert*! Es wäre offenbar verfehlt, hier überhaupt noch von *Entscheidungen* sprechen zu wollen. Was man sagen dürfte wäre lediglich, daß alle Leute, die sich zum Zeitpunkt der Untersuchung in einem gewissen (z.B. psychophysischen) Zustand $Z_1$ befanden, mit kausaler bzw. nomologischer Notwendigkeit später die Handlung $H_1$ *tun werden*, während jeder, der sich zu $t_U$ im konträren Zustand $Z_2$ befand, später zwangsläufig $H_2$ *tut*! Ein derart determiniertes Verhalten entzieht sich jedoch jeder Beurteilung als *rational* oder irrational.[10] Nach $t_U$ hat es niemand mehr in der Hand, sich so oder anders zu entscheiden. Dem Glücklichen, der sich zu $t_U$ im Zustand $Z_1$ befand, werden zu $t_G$ 1.000.000 $ in $K_2$ gelegt, und er kann danach seinen Gewinn weder aufs Spiel setzen noch um die 1.000 $ aus $K_1$ steigern, denn er vermag nichts anders zu tun, als zu $t_A$ $H_1$ zu „wählen". Wer sich dagegen zu $t_U$ im Zustand $Z_2$ befand, für den bleibt $K_2$ leer; der hat deshalb zwar keine Chance, durch eine spätere Entscheidung für $H_1$ noch Millionär zu werden, aber er läuft auch nicht Gefahr, den Trostpreis von 1.000 $ zu verlieren: Er kann nämlich nichts anderes tun, als zu $t_A$ $H_2$ zu „wählen".

---

8 Vgl. Olin [1976: 132]: „the states used as evidence [for the Being's predictions] are nomically linked to the actions. There is, let us suppose, a law which says that all and only those who do [$H_1$] are in state [$Z_1$] one hour before playing; similarly with regard to [$H_2$] and [$Z_2$]."

9 Man sollte meinen, daß jemand, der etwas so Komplexes wie die Entscheidungen von *Personen* vorherzusagen versteht, keinerlei Schwierigkeiten haben dürfte, so etwas Elementares wie das physikalisch determinierte Verhalten von 49 Kugeln in einem geschlossenen Behälter vorherzusagen, also die Resultate jeder Lotto-Ziehung vorherzuwissen. Das ist vielleicht die Erklärung dafür, wieso der Newcomb-Vorhersager so großzügig mit Millionen um sich schmeißen kann!

10 Vgl. auch Mackie [1977: 151], für den die Frage nach der Rationalität der Handlungen in diesem Fall „müßig" ist. Hingegen versuchen Gibbard/Harper [1978: 370/1] zu begründen, daß die Handlungsstrategie $H_2$ „even in the case where the predictor is known to be infallible" *rational* sei.

Der von manchen AutorInnen erteilte Ratschlag „Glaube dem Vorhersager, nimm bescheiden nur die Kiste $K_2$ und werde Millionär!“[11] erweist sich deshalb als *nutzlos*, sofern er dem Handelnden zum Zeitpunkt der Entscheidung, $t_E$, zu einem „rationalen“ oder gewinnmaximierenden Entschluß verhelfen soll. Zum *diesem* Zeitpunkt sind die Würfel längst gefallen! Eine entsprechende Empfehlung könnte höchstens dann Sinn machen, wenn man zu einer weiteren Modifikation der Situation übergeht, die allerdings mit dem ursprünglichen Newcomb-Problem nur noch wenig gemein hat.

---

[11] Vgl. z.B. Bar-Hillel/Margalit [1972: 303]: „[...] we hope to convince the reader to take just the one [...] box, and join the millionaire’s club!“

## 9 Intendieren und Handeln

Bei dieser u.a. in Gauthier [1988/89: 186 ff.] betrachteten Variante weiß der Spieler z von vornherein, daß er zum Zeitpunkt $t_U$ untersucht wird und daß sein jeweiliger Zustand $Z_1$ bzw. $Z_2$ dann den Ausschlag dafür gibt, ob anschließend die Situation $S_1$ oder $S_2$ realisiert wird. Dabei kann $Z_1$ bzw. $Z_2$ mit dem psychischen Zustand von z identifiziert werden, fest zu *beabsichtigen*, zu $t_E$ $H_1$ bzw. $H_2$ zu tun. Sofern man den „Akt" des Beabsichtigens überhaupt als eine Handlung betrachten darf, die einer Bewertung als rational oder irrational zugänglich ist, wird dann natürlich die „Metahandlung" $MH_1$, d.h. die Intention, später $H_1$ zu tun, *die rationale* Option. Die ursprüngliche Dominanz-Überlegung für $H_2$ läßt sich ja nicht auf die Metahandlung $MH_2$ übertragen, die darin bestehen würde, zu $t_U$ zu intendieren, zu $t_E$ $H_2$ zu tun, denn $MH_1$ bzw. $MH_2$ haben nach Voraussetzung einen entscheidenden *kausalen Einfluß* darauf, ob die 1.000.000 $ in die Kiste gelegt werden oder nicht. Hier lautet die richtige Empfehlung also tatsächlich: „Glaube dem Vorhersager, *intendiere*, bescheiden die Kiste $K_2$ zu nehmen, und werde Millionär!" So weit so gut!

Ein interessantes Nachfolge-Problem entsteht nun dann, wenn man annimmt, daß die späteren *Handlungen* $H_1$ bzw. $H_2$ von der früheren *Entscheidung* $MH_1$ (analytisch bzw. nomologisch) *unabhängig* sind. Könnte es nicht rational sein, *nach* der Untersuchung *gegen* die ursprüngliche Intention zu handeln? Verspricht also, mit anderen Worten, die „kombinierte Handlung" $MH_1 + H_2$ nicht einen noch größeren Gewinn als $MH_1 + H_1$? Nehmen wir an, daß 200 Personen bei einem solchen Spiel mitgemacht hätten, darunter 100 Gläubige $\{x_1,...,x_{100}\}$, die alle im Einklang mit ihrer Intention $MH_1$ brav $H_1$ taten, und ebenso viele Wagemutige $\{w_1,...,w_{100}\}$, die sich entgegen ihrer früheren Intention $MH_1$ für $H_2$ entschieden. Ferner sei empirisch verbürgt, daß wiederum 99 % der $x_i$ die 1.000.000 $ kassierten, während sich 99 % der $w_j$ mit 1.000 $ zufrieden geben mußten. Müßte man hier nicht wirklich urteilen, daß $H_1$ zu *tun* – und nicht bloß vorher zu *intendieren* – die rationale Entscheidung ist?

Nicht unbedingt! Die Existenz des einen gläubigen Pechvogels, x*, ebenso wie die des einen erfolgreichen Wagemutigen, w*, sollte eigentlich deutlich machen, daß das kombinierte Spiel nur die *Illusion* erzeugt, der Gewinn der 1.000.000 $ sei auf die *Handlung* $H_1$ zurückzuführen, während er in Wirklichkeit eine Folge der *Metahandlung* $MH_1$ darstellt. In völliger Analogie zu Abschnitt 6 gilt auch hier für beliebige $x_i \neq x^*$ und $w_j \neq w^*$:

(IK3') Wenn $x_i$ beide Kisten genommen hätte, dann hätte er 1.001.000 $ bekommen

(IK4') Wenn $w_j$ nur die Kiste $K_2$ gewählt hätte, dann hätte er gar nichts bekommen.

Spätestens durch Einnahme der Perspektive des wissenden Beobachters sollte einem wieder bewußt werden, daß *in der jeweiligen Situation*, in der sich die Spieler *zu* $t_E$ befanden, $H_2$ und nicht $H_1$ die gewinnmaximierende Alternative war. Allerdings müßten sich die 99 erfolglosen Wagemutigen nicht nur den hämischen Kommentar „If you're so smart, why ain't you rich?" gefallen lassen[12], sondern sie müßten sich nach der Replik „Weil ich es offenbar zu $t_U$ – trotz größter Anstrengung – nicht geschafft habe, die feste Absicht zu bilden, zu $t_E$ $H_1$ zu tun" auf die weitere Belehrung gefaßt machen, man könne eben nur dann die *volle* (echte, aufrichtige) Absicht haben, später $H_1$ zu tun, wenn man später tatsächlich $H_1$ tut.

Darin steckt sicher ein Körnchen Wahrheit. Vermutlich ist es nomologisch oder sogar analytisch notwendig, daß jemand nur dann wirklich *beabsichtigt*, später H zu tun, wenn er eine *Disposition*, H zu tun, entwickelt, die über einen gewissen Zeitraum hinweg so stark bleibt, daß er bei entsprechender Gelegenheit wirklich H tut[13]. Zumindest würde man zurecht an der Ernsthaftigkeit der Intention einer Person zweifeln, die von vornherein damit spekuliert, dieser ihrer Absicht später eventuell zuwider zu handeln. Deshalb stellt das Newcomb-Nachfolge-„Problem" vielleicht nur ein Pseudo-Problem dar, das auf einer unhaltbaren Prämisse hinsichtlich des (nomo-)logischen Verhältnisses von Intentionen und Handlungen beruht. Setzt man jedoch dem Argument zuliebe voraus, daß eine Abkoppelung der späteren Handlungen von den früheren Intentionen möglich ist, so bleibt ein weiteres interessantes Argument zu diskutieren, mit dem die Rationalität von $H_1$ begründet werden soll. Im Anschluß an Dennett [1986] betrachtet Leder [1996: 13 ff.]:

> „[...] den Tip des Golfprofis, beim Abschlag den Kopf unten zu lassen und den Schläger durchzuschwingen, nachdem man den Ball bereits getroffen hat. Ein vergleichbares Beispiel ist die Empfehlung, beim Top-Spin (im Tischtennis), nachdem man den Ball berührt hat, den Arm in ganz bestimmter Weise weiterzuführen [...]. Wie Dennett argumentiert und viele Sportler ohnehin wissen, hat dies mit abergläubischem Verhalten nichts zu tun. Etwa beim Top-Spin kann man nur, indem man nach der Ballberührung eine bestimmte Bewegung ausführt, sicherstellen, daß man während der Ballberührung das Richtige tut. Wir haben physisch keine Möglichkeit, die Bewegung bis zur Ballberührung genauso auszuführen und danach abzubrechen. Wenn wir sie abbrechen, ist sie auch vorher schon anders. Das heißt: Nur dadurch,

---

12 Vgl. Gibbard/Harper [1978: 371].

13 Mit Bezug auf das dem Newcomb-Problem verwandte „Toxin Puzzle" äußert Kafka [1983: 35] analoge Bedenken: „If intentions were inner performances or self-directed commands, you would have no trouble earning your million. [...] Similarly, if intentions were simply decisions, and decisions were volitions fully under the agent's control, there would be no problem. But intentions are better viewed as dispositions to act which are based on *reasons to act*".

> daß wir zu einem späteren Zeitpunkt etwas tun, was kausal völlig irrelevant ist, können wir sicherstellen, daß wir im entscheidenden Moment das Richtige tun.“

Leders Übertragung dieses Gedankens auf das Newcomb-Problem hinkt jedoch, weil beim kombinierten Spiel laut Voraussetzung (physisch und psychisch) durchaus die *Möglichkeit* besteht, zu $t_U$ optimal zu intendieren, zu $t_E$ $H_1$ zu tun, und später dennoch nicht $H_1$, sondern $H_2$ zu wählen: Nach Voraussetzung hat w* dies ja geschafft.

Daraufhin könnte Leder entgegnen, daß das Reden von physischer Unmöglichkeit nicht ganz wörtlich zu nehmen sei und daß eine hohe statistische Korrelation zwischen richtigem Top-Spin-Schlag und anschließendem „Durchziehen“ völlig ausreicht. Auch wenn irgendein Tischtennisspieler irgendwann einmal einen Top-Spin Ball optimal trifft, obwohl er die anschließende Bewegung abbricht, so wird man doch jedenfalls nur dadurch zu einem guten Top-Spin-*Spieler,* daß man das „Durchziehen“ regelmäßig trainiert. Entsprechend zeige ja auch der 99-%ige Mißerfolg der Wagemutigen, daß im kombinierten Newcomb-Spiel die Option $H_1$ (auf lange Sicht) die falsche *Strategie* ist. Dies werde spätestens beim Übergang zur nächsten Variante evident, bei der an Stelle der 100 Gläubigen und 100 Wagemutigen nur noch *ein einziger* Akteur im Spiel bleibt. Dieses *Super-Newcomb-Spiel* läuft so:

> Man nehme an, der Person x würde angeboten, eine Serie von 201 kombinierten Newcomb-Spielen gegen das mächtige Wesen W zu spielen. Bei jedem Durchgang soll x zunächst eine feste *Absicht* bilden, später $H_1$ bzw. $H_2$ zu tun. W unterzieht x einer psychologischen Untersuchung und deponiert – je nachdem ob er bei x die feste Absicht entdeckt hat, $H_1$ zu tun – die 1.000.000 in $K_2$ oder nicht. Danach trifft x ihre aktuelle *Entscheidung* und *tut* $H_1$ bzw. $H_2$.
>
> Zu Beginn jedes der mittlerweile 200 absolvierten Durchgänge versuchte x so gut wie nur irgend möglich, die Intention zu fassen, später $H_1$ zu tun; anschließend blieb x in insgesamt 100 Fällen dieser Intention treu, wählte jeweils $H_1$ und erzielte dabei 99-mal einen Gewinn von 1.000.000 $, während er nur ein einziges Mal mit leeren Händen dastand; entsprechend oft probierte x die $H_2$-Strategie und mußte sich in 99 Fällen mit 1.000 $ begnügen; ein einziges Mal erzielte er so den maximalen Gewinn von 1.001.000 $. Wäre es aufgrund dieser Erfahrungen (und sofern x immer noch daran interessiert ist, eher 1.000.000 als nur 1.000 $ zu bekommen) nicht vollkommen rational, beim Schlußversuch $H_1$ zu tun?

Diese exotische Variante des Newcomb-Nachfolge-Problems vermag ich (zumindest in diesem Rahmen) nicht mehr zu lösen. Vielleicht kommt in Zukunft ein anderer Kollege auf den Gedanken, in einer anderen Festschrift zu einem anderen runden Geburtstag dieses Problem weiter zu verfolgen.

## Literatur

BAR-HILLEL, Maya & MARGALIT, Avishai [1972]: „Newcomb's Paradox Revisited“. *British Journal for the Philosophy of Science* 23, 295-304.

DACEY, Raymond et. al. [1977]: „A Cognitivist Solution to Newcomb's Problem“. *American Philosophical Quarterly* 14, 79-84.

DENNETT, Daniel C. [1986]: *Ellbogenfreiheit – Die wünschenswerten Formen von freiem Willen*. Meisenheim (A. Hain).

EELLS, Ellery [1985]: „Causality, Decision, and Newcomb's Paradox“. In R. Campbell & L. Sowden (eds.), *Paradoxes of Rationality and Cooperation – Prisoner's Dilemma and Newcomb's Problem*. Vancouver (University of British Columbia Press), 183-213.

GÄRDENFORS, Peter & SAHLIN, Nils-Eric (eds.) [1988]: *Decision, Probability, and Utility. Selected readings*, Cambridge 1988.

GAUTHIER, David [1988/89]: „In the Neighbourhood of the Newcomb-Predictor. (Reflections on Rationality)“. *Proceedings of the Aristotelian Society* 89, 179-194.

GIBBARD, Allan & HARPER, William L. [1978]: „Two Kinds of Expected Utility“. In C.A.Hooker, J.J. Leach & E.F. McClennen (eds.), *Foundations and Applications of Decision Theory*. Vol 1, Dordrecht (Reidel), 125-162. Hier zitiert nach dem Abdruck in Gärdenfors/Sahlin [1988], 341-376.

GOODMAN, Nelson [1955]: *Fact, Fiction, and Forecast*. Cambridge, MA (Harvard University Press).

HORGAN, Terence [1981]: „Counterfactuals and Newcomb's Problem“. *Journal of Philosophy* 78, 331-356.

KAFKA, Gregory S. [1983]: „The Toxin Puzzle“. *Analysis* 43, 33-36.

KUTSCHERA, Franz von [1973]: *Einführung in die Logik der Normen, Werte und Entscheidungen*. Freiburg (Alber).

KUTSCHERA, Franz von [1974]: „Indicative Conditionals“ *Theoretical Linguistics* 1, 257-269.

KUTSCHERA, Franz von [1976]: *Einführung in die intensionale Semantik*. Berlin (de Gruyter).

LEDER, Matthias [1996]: *Was heißt es, eine Person zu sein?* Dissertation FU Berlin; hier zitiert nach dem unveröffentlichten Manuskript.

LEWIS, David [1973]: *Counterfactuals*, Oxford (Blackwell).

LEWIS, David [1979]: „Counterfactual Dependence and Time's Arrow“. *Noûs* 4, 455-476.

LEWIS, David [1981]: „Causal Decision Theory“. Abgedruckt in Gärdenfors/Sahlin [1988], 377-405.

MACKIE, John L. [1977]: „Newcomb's Paradox and the Direction of Causation“. *Canadian Journal of Philosophy* 7, 213-225; hier zitiert nach dem Abdruck in J.L. Mackie, *Logic and Knowledge. Selected Papers Vol. 1*, Oxford 1985, 145-158.

NOZICK, Robert [1969]: „Newcomb's Problem and Two Principles of Choice“. In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*, Dordrecht (Reidel), 114-146.

OLIN, Doris [1976]: „Newcomb's Problem: Further Investigations“. *American Philosophical Quarterly* 13, 129-133.

SPOHN, Wolfgang [1978]: *Grundlagen der Entscheidungstheorie*. Kronberg/Ts. (Scriptor).

STALNAKER, Robert [1968]: „A Theory of Conditionals“. In N. Rescher (ed.), *Studies in Logical Theory*, Oxford (Blackwell), 98-112.