

NewComb'sche Paradoxie:

Ein Wesen W, das mit sehr hoher Wahrscheinlichkeit p das Verhalten (Entscheidung) einer Person A vorhersagen kann, erläutert A folgendes Spiel:

W besitzt zwei Kästchen, deren Inhalt A nicht sehen kann. In dem linken Kästchen K1 liegen bereits 1000 €.

Das rechte Kästchen K2 wird W unter einer noch zu nennenden Bedingung B mit einer Million Euro füllen.

A darf nun zwischen zwei Handlungen wählen:

H1: A öffnet beide Kästchen (und erhält das in ihnen befindliche Geld.)

H2: A öffnet nur das rechte Kästchen K2 (und erhält das in ihm befindliche Geld.)

1. Schritt: W prognostiziert (ohne, daß A die Prognose von W erfährt) das Wahlverhalten von A.

2. Schritt: W handelt nun nach folgenden Bedingungen:

Wenn $\underbrace{\text{W voraussieht, daß A beide Kästchen öffnet (H1)}}_{\text{Bedingung B1}}$, dann $\underbrace{\text{legt es nichts in K2.}}_{\text{F1}}$.

$\underbrace{\text{Sieht es voraus, daß A nur das rechte öffnet (H2)}}_{\text{Bedingung B2}}$, so $\underbrace{\text{legt es in K2 die Million Euro.}}_{\text{F2}}$.

Wie gesagt, W handelt nun danach.

3. Schritt: W teilt diese Bedingungen nun A mit.

4. Schritt: A ist an der Reihe und entscheidet sich für H1 bzw. H2.

Die Strategie von W ist also: $\text{Prog}_W(\text{H1}) \Rightarrow \text{F1}$
 $\text{Prog}_W(\text{H2}) \Rightarrow \text{F2}$

Zwei mögliche Strategien für A:

1.Strategie: Wenn A beide Kästchen öffnet, dann läßt W K2 leer (da es H1 ja vorhersieht).
Ergebnis:

A erhält 1000 €.

Wenn A nur das rechte Kästchen öffnet, so wird W eine Million in K2 legen.
A erhält dann eine Million.

Also muß A nur das zweite Kästchen öffnen.

2. Strategie: Betrachtet man die Zeitfolge der Schritte, so hat W sich schon entschieden und das Kästchen zwei mit einer Million gefüllt oder nicht, je nach der Prognose von A's Verhalten.

Wenn nun die Million im 2. Kästchen liegt, und A sich für das Öffnen beider Kästchen entscheidet, so gewinnt er 1001000 €.

Wenn Kästchen zwei aber leer ist und A beide Kästchen öffnet, erhält er wenigstens 1000 €.

In beiden Fällen hat A also 1000 € mehr als bei der anderen Entscheidung.
A wird also beide Kästchen öffnen.

Nehmen wir an, daß W mit Sicherheit eine der beiden Prognosen aufstellt.

A müßte dann folgende Überlegung Ü anstellen:

$$\text{Prog}_W(t_1, H_1) \Rightarrow \text{K2 leer} \xrightarrow{p} H_1(t_2) \Rightarrow \text{Euro} = 10^3$$

$$\text{Prog}_W(t_1, H_1) \Rightarrow \text{K2 leer} \xrightarrow{1-p} H_2(t_2) \Rightarrow \text{Euro} = 0$$

$$\text{Prog}_W(t_1, H_2) \Rightarrow \text{K2 1 Million} \xrightarrow{p} H_2(t_2) \Rightarrow \text{Euro} = 10^6$$

$$\text{Prog}_W(t_1, H_2) \Rightarrow \text{K2 1 Million} \xrightarrow{1-p} H_1(t_2) \Rightarrow \text{Euro} = 10^6 + 10^3$$

Wie sollte A nun im eigenen Interesse handeln? Hat er überhaupt noch Handlungsmöglichkeiten, wenn die Voraussetzung der Prognosefähigkeit von W stimmt.

W muß annehmen, daß A max. Geld will. Also wird A im Rahmen seiner Möglichkeiten entsprechend handeln, wenn er vernünftig handelt.

Dabei stellt sich natürlich die Frage nach der Bedeutung der Prognoserichtigkeit von W. In welcher Hinsicht stellt W seine Prognosen. Wie muß man den Handlungsverlauf verstehen. Ist er ein vernünftiger, der nach den geeigneten Mitteln Ausschau hält und halten kann oder ist er ein quasikausaler? Wenn W Prognosen stellen kann, so auf der Grundlage von zwei Möglichkeiten. Man nimmt an (Th1), daß alles nach NaturGesetzen verläuft, die man im Sinne von Verlaufstendenzen interpretiert. Oder man unterstellt, daß Handeln vernünftigen Regeln unterliegt, die durch Einsicht in Mittel-Zweck-Beziehungen aufgestellt werden (Th2).

Ist es überhaupt möglich von Verlaufstendenzen zu sprechen, wenn man keine Zweck-Mittel-Überlegungen unterstellt? Kann ich solche Tendenzen feststellen ohne irgend ein Interesse zu haben?

Eine Tendenz besagt doch, daß in Situationen, die ich als A-Situationen feststelle, bemerke, interpretiere, mit einer gewissen rel. Häufigkeit h B-Situationen auftreten, sodaß ich sage, daß die A-Situationen die Tendenz h zu B-Situationen haben. Dazu gehört, daß ich solche Situationen best. Art (A oder B) bestimmen kann. Hätte ich kein Interesse (pos oder neg.) an A jetzt oder früher gehabt, so würde ich es nicht als A bestimmen können.

Wäre es nicht auf der anderen Seite annähernd 'tatsächlich' eine Situation, die ich mit A bezeichnet kann aufgrund ihrer besonderen Beschaffenheit, so wären diese Beschreibungen eben auch nicht möglich. Zu einer A-Situation gehören also best. 'objektive' Sprödigkeiten und subjektives Interesse. Nur durch beides gibt es A-Situationen. Weiter: Gäbe es kein Interesse, was mich bewegt 'Tendenzen' zu bemerken und sie wiederum festzustellen, dann gäbe es sie auch nicht. Ebenso wenig gäbe es sie, wenn nicht eine Sprödigkeit gegenüber bel. Möglichkeiten existieren würde und alles möglich wäre. Ich muß sie 'objektiv' bemerken können. Es muß etwas sein, daß ich eben sage es verändere sich so und nicht ganz anders. Illusionen allein können es nicht sein.

Interessen und Sprödigkeit können nicht im gleichen Sinn wieder Gegenstände von Interesse und Objektivität sein wie ihre Gegenstände, will man nicht in einen Regress ad infinitum riskieren.

Könnte niemand aktiv Veränderungen bewirken, ginge alles seinen 'natürlichen Gang', von A zu B etc, den ich nur feststellen würde, so fragt sich wieder, wie das möglich sein soll. Wie entstünden meine Interessen?

Ist ein Interesse möglich, das rein 'natürlich' befriedigt wird? Dann wäre es sicher gänzlich sinnlos. Sein Dasein oder NichtDasein würde keinen Unterschied machen. Die Resultate wären die gleichen. Die Existenz eines Interesses setzt notwendig voraus, daß sie etwas bei Befriedigung verändern. Und daß die Veränderung aktiv - von mir oder jemandem anderem herbeigeführt wird. Und gäbe es keine Interessen, so gäbe es überhaupt nichts Bestimmtes. Kein A, kein B und somit keine Veränderung...

Also sieht es so aus, als ob es aktive Veränderung geben müsse, damit es überhaupt Veränderung geben kann.

Aktive Veränderung gibt es aber nur, wenn es Zwecke oder Ziele gibt oder sie gesetzt sind (von mir und/oder anderen). Tendenzen gibt es also nur, wenn es Zwecke gibt. Verlaufstendenzen sind also ohne Zwecke grundsätzlich nicht denkbar. Prognosen über Verlaufstendenzen benötigen daher (zumindest fremde) Zwecke und mit ihnen aktive Handlungen, die nicht selbst wieder (grundsätzlich) als Prozesse interpretierbar sind.

Reine Prognosen über Naturphänomene sind also sinnlos.

Zumindest das interpretierende Wesen muß also bei Prognosenaufstellung ein zweckvoll Handelndes sein.

Damit ist These 1 widerlegt.

Was sind vernünftige Regeln?

Handlungsmittel, die bei Anwendung und Einwirkung auf die Welt eine A-Situation in eine B-Situation verwandeln, wobei eine B-Situation bzw. Folgen davon intendiert sind. Daß eine A-Situation in eine B-Situation verwandelt wird, ist wie gesagt kein pures Faktum der Natur. (Sonst wäre die Verwandlung nicht prognostizierbar.)

Eine Entscheidung, Handlung, kann ich prognostizieren in dem Sinn, daß ich bei Unterstellung der Vernünftigkeit, bei Unterstellung der Urteilsfähigkeit und der Einsicht in

Prozesstendenzen, ich annehme, daß bei ähnlichen Situationen, (A-Situationen), die Person die B-Situation herstellen wird, wenn sie die beste ist zur Erreichung des Zwecks. Das hängt natürlich von vielen Faktoren ab: Gesundheit, Glauben etc..

(Eine ganz andere Möglichkeit: Es gibt verschiedene parallele Räume verschiedener Zeiten (Kleinsche Flasche). Prognose wäre der Fernblick in einen Nebenraum, der für uns die Zukunft bedeutete.)

Also:

W muß also annehmen, daß A die Überlegung Ü anstellt. A würde natürlich versuchen, möglichst viel Geld zu erhalten. Das meiste bekommt er, wenn W die Entscheidung H2 vorhersagt. W wird diese aber dann (zunächst) vorhersagen, wenn er meint, A glaube, er gewänne bei dieser Entscheidung den größten Betrag (A handelt ja rational). Das tut er aber nicht. Denn den größten Betrag erhält A, wenn er sich für H1 entscheidet und W glaubt, er würde sich für H2 entscheiden. Das kann W aber nicht, da er die Überlegung ja auch anstellt und gegen sein Wissen kann er nicht das Gegenteil glauben. Und A weiß das auch. Also wird A nicht H1 tun, sondern H2.

Das weiß auch W. Daher prognostiziert er H2. (*)

A wird das ebenfalls auch wissen und daher wieder H1 überlegen zu tun und sein Ziel erreichen....wenn das nicht eben auch W voraussieht.

W wird also H1 prognostizieren müssen. Würde daraufhin A H2 tun, so würde er nichts erhalten. Also tut er H1. Das weiß auch W, daher prognostiziert er H1 und A tut H1 und gewinnt leider nur 1000 €.

Diese Strategie ist für A aber schlecht und W weiß das. Daher wird A, wenn W H2 prognostiziert (*), nicht H1 tun, sondern H2 und erhält eine Million. Das weiß auch W. usw. und so fort wieder nach (*) mit einer neuen Möglichkeit wieder H1 zu probieren.

Wie weit soll A denken und wie weit wird er denken. Es hängt eben davon ab, wie weit W denkt und das kann A nicht wissen und auch W weiß nicht wie weit A denkt.

Es bleibt also dem Zufall überlassen, was W prognostiziert und was A tut. Das wissen beide. Das bedeutet, daß die Entscheidung von A nicht mehr mit p bzw. 1-p aus der Prognose von W folgen kann, es sei denn p ist 0,5. Die Prognosen selbst sind dann auch 50 zu 50.

A darf also einen Gewinn von $(1001000/2) = 500500$ € rechnen.

$$E = \frac{1}{4} \cdot 1000 + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 10^6 + \frac{1}{4} \cdot (10^6 + 10^3) = \frac{1}{2}(1001000) = 500500$$

