

Statistiken, Modellbildung, Methoden, Management

Version 1.0, © Oktober 2007, Guido Stepken

Vielen Entscheidungen - darunter auch mitunter sehr gewichtigen - liegen Statistiken zugrunde. Sie basieren auf Modellen und dienen gleichzeitig auch wieder der „Modellbildung“, bzw. Korrektur von Modellen, sollen Vorurteile, Annahmen, Überzeugungen bestätigen bzw. widerlegen. Ist einmal ein faktizitätstreues, wirklichkeitsgerechtes Modell gefunden, so entscheidet das Methodenwissen der Entscheider und Handelnden über Erfolg und Misserfolg. Unternehmen, Organisationen sind Entscheidungssysteme, welche auf Ergebnissen komplexer Untersuchungen, Analysen und Metriken beruhen, welche mit bestimmten Methoden gewonnen wurden. Ob nun bei Wahlen, bei Nachweis der Schädlichkeit von Schadstoffen, Medikamenten, bei Umfragen, bei der Beurteilung von psychologischen Tests, u.s.w. - überall besteht die Möglichkeit, die Ergebnisse völlig anders darzustellen, als sie eigentlich gemessen wurden. Das Interessante dabei - Jeder Schritt lässt sich wissenschaftlich exakt nachvollziehen und die Aussage ist noch nicht einmal gelogen. Angeblich „bewiesene“ Scheinkorrelationen führen darüber hinaus noch zu Überzeugungen, welchen völlig falsche Weltbilder bzw. „mentale Modelle“ zugrunde liegen. Wie dies kann, sei nun hier in meinem Essay ausführlich erklärt und vorgerechnet, also „bewiesen“!

Aggregationsfehler oder Simpsons Paradox

Zunächst einmal sei ein interessanter Effekt vorgestellt, welcher **Aggregationsfehler** oder auch **Simpsons Paradox** genannt wird, und mit welchem sich in der Statistik alles beweisen lässt, insbesondere bei Tierversuchen und Schädlichkeiten von Medikamenten, Nahrungsmittel-Zusatzstoffen. Wer es versteht, damit zu jonglieren, kann Statistik-Gläubigen, dazu gehören u.a. auch fast alle Entscheider in Führungspositionen, denen dieser Effekt oft unbekannt ist, und welche leider oft Statistiken als Begründung ihrer oft millionen - und milliardenschweren Entscheidungen heranziehen bzw. anfertigen lassen, wie Ingenieure, Wissenschaftler, Psychologen, Politiker

Hierzu zuerst eine Statistik bezüglich zweier verschiedener Schadstoffen A und B, in 2 Versuchsreihen an Ratten und Mäusen getestet:

| Schadstoff A | Anzahl | Tote | Rate |
|---------------------|---------|------|--------|
| Maus | 100000 | 150 | 0,015 |
| Ratte | 30000 | 150 | 0,050 |
| | | | |
| Schadstoff B | Anzahl | Tote | Rate |
| Maus | 4000000 | 8000 | 0,020 |
| Ratte | 100000 | 1000 | 0,010 |
| | | | |
| Nagetiere | Gesamt | Tote | Rate |
| Schadstoff A | 130000 | 300 | 0,0231 |
| Schadstoff B | 4100000 | 9000 | 0,0220 |

Daraus lassen sich nun folgende Interpretationen ableiten:

1. Aussage: „Die Sterblichkeit von Mäusen ist bei Schadstoff A um 25% geringer!“
2. Aussage: „Die Sterblichkeit von Ratten ist bei Schadstoff A sogar um 50% geringer!“

Und nun kommt der Profi-Statistiker und verdreht durch seine durchaus an Zahlenwerten „beweisbaren“ Statistiken völlig ins Gegenteil: „Die Sterblichkeit bei Nagetieren ist bei Schadstoff B geringer!“

Wohlmerkt: Grundlage ist dasselbe Datenmaterial!

Bewertung zweier Eignungstests zur Einstellung von Bewerbern

Es wurden zwei verschiedene Verfahren (A und B) zur Bewertung von Kandidaten bei Psychometrischen Einstellungstests evaluiert. Hierbei wurden in zwei Versuchsreihen über mehrere Jahre in Unternehmen der Werdegang von weit über 400.000 Bewerbern auf Führungspositionen ausgewertet.

| Verfahren A | Anzahl | Fehler | Fehlerquote |
|--------------------|--------|--------|-------------|
| Versuchsreihe 1 | 10000 | 150 | 0,0015 |
| Versuchsreihe 2 | 3000 | 15 | 0,050 |
| Verfahren B | | | |
| | Anzahl | Fehler | Fehlerquote |
| Versuchsreihe 1 | 400000 | 800 | 0,02 |
| Versuchsreihe 2 | 10000 | 100 | 0,01 |
| Gesamt | | | |
| | Anzahl | Fehler | Fehlerquote |
| Verfahren A | 13000 | 30 | 0,0231 |
| Verfahren B | 410000 | 900 | 0,0220 |

Daraus lassen sich nun folgende Interpretationen ableiten:

1. Aussage: „In Versuchsreihe 1 hat Verfahren A zu 25% weniger Fehlentscheidungen geführt, und daher signifikant besser abgeschnitten!“
2. Aussage: „In Versuchsreihe 2 hat Verfahren A sogar zu zu 50% weniger Fehlentscheidungen geführt, und daher signifikant besser abgeschnitten!“

Diese Aussagen zeigen auf, dass Firma A das offensichtlich bessere und zuverlässigere Verfahren für Eignungstests von Bewerbern in Unternehmen besitzt.

Und nun kommt der Statistiker und verdreht durch seine durchaus an Zahlenwerten „beweisbaren“ Statistiken völlig ins Gegenteil, indem er beide Versuchsreihen für Verfahren A und B zusammenzieht, also „aggregiert“, und hierbei kommt dann eine völlig gegenteilige Aussage zustande:

Aussage: „Die Fehlerquote bei Verfahren B ist um 5% geringer!“

Wohlgemerkt: Auch hier gilt: Grundlage ist dasselbe Datenmaterial!

Beide Statistiken ähneln sich. Interessant hierbei ist, warum genau sich plötzlich eine qualitative Aussage ins Gegenteil verkehren kann, ohne dass Quantitäten sich ändern. Obiger Trick funktioniert nämlich nur dann, wenn zwei Verfahren erstens nicht auf derselben Menge beruhen, und zweitens wenn keine „Gleichverteilung“ zuvor nachgewiesen wurde.

Bewertung zweier Medikamente

Es wurden zwei Medikamente in Versuchsreihen mit Männern und Frauen miteinander verglichen:

| Weiblich | | | | |
|----------------------|----------|------|-------|---------------|
| Medikament | Genesung | | Summe | Genesungsrate |
| | Ja | Nein | | |
| neu | 18 | 12 | 30 | 60% |
| alt | 7 | 3 | 10 | 70% |
| Männlich | | | | |
| Medikament | Genesung | | Summe | Genesungsrate |
| | Ja | Nein | | |
| neu | 2 | 8 | 10 | 20% |
| alt | 9 | 21 | 30 | 30% |
| Gesamttabelle | | | | |
| Medikament | Genesung | | Summe | Genesungsrate |
| | Ja | Nein | | |
| neu | 20 | 20 | 40 | 50% |
| alt | 16 | 24 | 40 | 40% |

Auch hier kann man den typischen Verlauf einer „Aggregation“ sehen. Weibliche und männliche Testkandidaten wurden zusammengezogen. Obwohl das alte Medikament bei Frauen und Männern besser beurteilt wurde, hat sich das Verhältnis bei Beurteilung des Medikamentes bei „Mensch“ - also weder männlich und weiblich - plötzlich völlig verkehrt. Das neue Medikament war plötzlich besser.

Ein Ding der statistischen Unmöglichkeit? Nein, einfach ein falsches Prinzip, welches fast allen Wissenschaftlern heute noch, also nach Jahrzehnten der allgemeinen Aufdeckung, anwenden. Nicht nur das Datenmaterial hier ist völlig unzureichend, was man sehr häufig bei Diplomarbeiten in der Psychologie sehen kann. Auch werden oft Theorien auf völlig unzulässige Weise aus anderen Wissensgebieten übernommen, und wo versucht wird, Korrelation auf der Ebene der Statistik nachzuweisen, ohne das hier logisch und kausal ein Zusammenhang bestünde. Dieses Phänomen habe ich „Metabasis“ genannt. *Siehe auch Foppa 1986, S.151 über das „Traditionelle Induktionsproblem“: „Wieviele Fälle, Variablen und Situationen man auch untersucht, meistens handelt es sich nur um einen kleinen Teil sämtlicher Möglichkeiten. Das hat zur Folge, dass man nicht von der allgemeinen Geltung dessen ausgehen kann, was man festgestellt hat, sondern die Geltung von Fall zu Fall prüfen muss“.*

Inverse Aggregationsfehler

Obiges Verfahren funktioniert prima auch umgekehrt, weswegen ich dieses Verfahren auch „Inversen Aggregationsfehler“ genannt habe. In der Literatur ist er nicht bekannt, jedoch leicht nachzuvollziehen:

| | Anzahl | Fehler | Quote |
|--------------------|---------------|---------------|--------------|
| Verfahren A | 10000 | 300 | 0.0300 |
| Verfahren B | 10000 | 300 | 0.0300 |

Irgendwie kommen - sagen wir mal bei einer Diplomarbeit - die Ergebnisse nicht wie gewünscht heraus. Beide Verfahren bei gleicher Normbasis scheinbar völlig gleich. Nun, da kann einfach nachgeholfen werden, indem man einfach zwei Messreihen „erfindet“. Erstens kann es sowieso niemand wirklich nachprüfen, und ausserdem liegen die Messwerte alle noch locker im Rahmen des möglichen bei der „Gleichverteilung“. Ich splitte die Messungen über 10000 in zwei „asymetrische Versuchsreihen“ einfach auf: 9000+1000 und 1000+9000. Die Fehler lasse ich von der absoluten Zahl her einmal gleich und das zweite mal passe ich die Summen nur noch an, damit beim Zusammenziehen 300 herauskommt:

| Verfahren A | Anzahl | Fehler | Quote |
|--------------------|---------------|---------------|--------------|
| Versuchsreihe 1 | 9000 | 150 | 0.0167 |
| Versuchsreihe 2 | 1000 | 150 | 0.1500 |
| | | | |
| Verfahren B | Anzahl | Fehler | Quote |
| Versuchsreihe 1 | 1000 | 25 | 0.0250 |
| Versuchsreihe 2 | 9000 | 275 | 0.0306 |

Hehe! Nun habe ich schon „auseinanderdriftende Werte“, jedoch irgendwie passt das alles noch nicht:

In Versuchsreihe 1 war Verfahren A mit ca. 30% weniger Fehlern behaftet (rot)

In Versuchsreihe 2 war Verfahren A dafür mit 5x so vielen Fehlern behaftet (grün)

Das widerspricht sich. Klarer wird hierdurch meine Statistik nicht ;-(

Also muss ich noch einige weitere Versuche durchführen, um meine Statistik „signifikant zu erhärten“.

Den oberen Teil verändere ich nur ein wenig, den unteren Teil „tune“ ich so richtig:

| Verfahren A | Anzahl | Fehler | Quote |
|--------------------|---------------|---------------|--------------|
| Versuchsreihe 1 | 8000 | 150 | 0.0167 |
| Versuchsreihe 2 | 2000 | 150 | 0.0750 |
| | | | |
| Verfahren B | Anzahl | Fehler | Quote |
| Versuchsreihe 1 | 41000 | 780 | 0.0190 |
| Versuchsreihe 2 | 9000 | 720 | 0.0800 |

Und schon passt wieder alles: Ich habe bei Verfahren B mal eben die 10-fache Anzahl an Tests gemacht, ich war halt „fleissig“. Der Rest ist geringfügig angepasst. Und nun ist in beiden Versuchsreihen das Verfahren A besser! Und die zusammengezogene Quote stimmt ja nun auch wieder. Ist also nicht wirklich gelogen gewesen, sondern ich habe die Zahlenbasis bei Verfahren B nur halt proportional etwas „verlängert“, „erweitert“, „vermehrte“ verfünffacht ...

| | Anzahl | Fehler | Quote |
|--------------------|---------------|---------------|--------------|
| Verfahren A | 10000 | 300 | 0.0300 |
| Verfahren B | 50000 | 1500 | 0.0300 |

Eine ausführliche Analyse der Gleichverteilung ist also ein wesentliches Kriterium bei Statistiken. Misstrauisch muss man als Gutachter immer dann werden, wenn jemand ein Verfahren ausführlicher, als das andere getestet hat. Allgemeine Sichtweise ist, dass Verfahren B durch die grössere Zahlenbasis „erhärter“ sei. Auch weichen die Werte bei Versuchsreihe A und B kaum voneinander ab, weswegen hier auch nichts auffällig erscheint: 0.0167 zu 0.0190 und 0.075 zu 0.080. Alles im Rahmen normaler Schwankungen! Aber die haben es in sich ;-)

Noch verrücktere Aggregationsfehler

Viel mehr Unsinn kann man durch Unterscheidungen in noch mehr Verfahren und Messreihen anstellen:

| <i>Alter zu Beginn der Studie</i> | | <i>Verstorben</i> | <i>Lebend</i> | <i>Summe</i> | <i>Quotient</i> |
|-----------------------------------|--------------|-------------------|---------------|--------------|-----------------|
| 55-74 | Raucher | 80 | 71 | 151 | 47% |
| | Nichtraucher | 141 | 109 | 250 | 44% |
| 55-64 | Raucher | 51 | 64 | 115 | 56% |
| | Nichtraucher | 40 | 81 | 121 | 67% |
| 65-74 | Raucher | 29 | 7 | 36 | 20% |
| | Nichtraucher | 101 | 28 | 129 | 22% |

Aussage: Raucher leben länger, als Nichtraucher, ausser, sie sind jünger oder älter als 65 Jahre!

Problem: Im oberen Teil der Tabelle hat man viele junge Nichtraucher mit alten Nichtrauchern vermischt, und nachgeschaut, wer von beiden länger lebt. Die Ungleichverteilung ergibt sich quasi erzwungen von alleine hierdurch und verfälscht die Statistik. Logikfehler durch und durch, obwohl alles – auf den ersten Blick nach dem „gesunden Menschenverstand“ beurteilt, „ok“ aussieht.

Siehe auch: Hans-Peter Beck Bornholdt und Hans-Herrmann Dubben:

Der Hund, der Eier legt - Erkennen von Fehlinformation durch Querdenken

Suggestivanteile in Umfragen und „zyklisches“ Hinterfragen

Ein weiteres Problem ist der Suggestivanteil in der Fragestellung, siehe Gerald Zaltman: „How customers think!“. Er stellte fest, dass eine Umfrage zur Kundenzufriedenheit bei Daimler-Benz die falschen Resultate zeigte: „Wie zufrieden sind Sie mit dem Kundenservice?“, etc. Er bemängelte, dass die Art der Fragestellung den Kunden gedanklich auf eine ja – Schiene schickt, an dessen Ende doch ein positives Bild steht. Fragte man anders herum: „Welche schlechten Erfahrungen haben Sie mit dem Mercedes – Kundenservice, der Qualität von Mercedes Produkten gemacht?“ schickt den Kunden gedanklich auf die Suche nach den Fehlern und Pannen. Eine Negativ-Suggestive Frage traute sich eine Marketing – Abteilung von Daimler-Benz nicht zu stellen. Es kann nicht sein, was nicht sein darf! Und Corporate Identity duldet halt keine Nestbeschmutzung – Psychodynamisches Prozesse. Gerald „Gerry“ Zaltman fragte anders: „Was denken Sie, was Mercedes Benz über Sie denkt?“ Und das Ergebnis: „Dukatenesel“, „Muss Mercedes aus Imagegründen fahren, Firmenwagen“, „Verträge ...“, ... Gerry bescheinigte Mercedes ein gestörtes Verhältnis von Kunde zu Firma. Und er behielt Recht. Platz 37 der ADAC-Pannenstatistik bewies es, bis der Brainstorming – Abteilung bei Daimler-Chrysler die rettende Lösung einfiel. Sie informierten alle Mercedes – Fahrer, dass sie im Falle einer Panne nicht mehr die „gelben Engel“, sondern direkt den neu eingeführten Pannendienst von Mercedes anrufen sollten. Daimler-Benz rückte in der Pannenstatistik viele Stellen nach oben. Tatsächlich jedoch lassen Firmeninhaber ganze Flotten von gerade neu geleasteten Mercedes zurückgehen und nehmen stattdessen Audi oder Toyota.

Das Erkennen von Korrelationen

Statistische Zusammenhänge zwischen Ereignissen gehören zu den wichtigsten empirischen Hinweisen auf zugrunde liegende, mögliche „Kausalzusammenhänge“. Auch hier sind „inverse Aggregationsfehler“ möglich. Nun, in den mentalen Modellen der Physik liegt das Prinzip „Ursache und Wirkung“ zugrunde, das „Newton'sche“ Prinzip von Kraft als Ursache und der Beschleunigung als Wirkung. Alltäglich erfahrbar, und daher allgemein als „richtig“ beurteilt. Auch stimmt dieses Modell wunderbar mit dem religiösen Modell der Religion mit „Gott als erstem Beweger“ überein. Kein vernünftiger Mensch würde angesichts dieser überwältigenden, auch sinnlich erfahrbaren „Fakten“ an diesem Prinzip zweifeln. Dennoch sind in vielen Fällen rekursive Modelle, wo die Wirkung wieder zur Ursache wird, treffender. Was wir sinnlich erfahren können, dem wird primär Glauben geschenkt, und diesem „Wissen“ - nämlich dem der eigenen Erfahrung - wird stets der Vorzug gegeben. Und wir bilden uns aus der sinnlich wahrnehmbaren Erfahrung unsere mentalen Modelle:

◆

A child
does not dis-
cover the world by
learning abstract rules.
Instead it learns by looking
at concrete examples. An example
contains the rules as well. In contrast to
rules, the recognition of examples can be bas-
ed on tangible reality. The knowledge extracted from
an example serves as a **PATTERN** that is used to remember
facts and to construct new solutions. When grown-ups
are about to learn something or have to apply
unknown tools, they are put into a child's
position again. They will favor concrete
examples over abstract rules. The
rules will happily be gen-
erated automatically,
for this is how
the brain
works.

◆

Deshalb ist die Frage interessant, wie gut wir im Alltag darin sind, solche Zusammenhänge zu erkennen. Experimente haben diese Fähigkeit untersucht und eine Reihe interessanter Erkenntnisse gewonnen. Smedslund hat in einem bereits 1963 durchgeführten Experiment einige Krankenschwestern gebeten, 100 Karteikarten durchzusehen, die jeweils Informationen darüber enthielten, ob ein bestimmter Patient ein Symptom zeigte oder nicht und ob der Patient an einer Krankheit litt oder nicht:

| | | Krankheit | |
|----------------|----------|------------------|----------|
| | | anwesend | abwesend |
| Symptom | anwesend | 37 | 33 |
| | abwesend | 17 | 13 |

Tabelle: Kontingenztafel zu einem Zusammenhang zwischen einem Symptom und einer fiktiven Krankheit (Smedslund, 1963).

Die Tabelle zeigt die Statistik, die den Probandinnen vorgelegt wurde. So fand sich etwa bei 37 Patienten, die die Krankheit hatten, auch das Symptom. Die Tabelle zeigt Daten, die im Wesentlichen dafür sprechen, dass zwischen Krankheit und Symptom kein Zusammenhang besteht. Bei der Mehrheit der Phänomene (37) waren ja Symptom und Krankheit gleichermaßen anwesend wie auch abwesend. Gutgläubige, leichtgläubige und insgesamt von Charakter wohlwollende, also empathische Menschen achten viel mehr auf eine positive Bestätigung, dass hier ein offensichtlicher Zusammenhang besteht. Die vom Charakter her eher „kritischen Hinterfrager“ jedenfalls prüfen immer und überall die sog. „Umkehrlogiken“: „Besteht die Krankheit in etwa genauso häufig, ohne dass sich Symptome zeigen?“ Ja, so ist es, wie die Tabelle ebenfalls zeigt! Und zwar ist die Krankheit ebenso häufig auftretend, wenn das Symptom anwesend ist (53%), wie wenn es abwesend ist (57%).

Und nun kommt die Überraschung: 85 Prozent der Krankenschwestern an, dass ein Zusammenhang zwischen Symptom und Krankheit bestehe. Obgleich solche Verzerrungen nicht unter allen Lernbedingungen zu beobachten sind (vgl. Wasserman, 1990), zeigt doch eine große Zahl von Studien, dass es sich um ein allgemeines, häufig zu beobachtendes Phänomen handelt (Arkes & Harkness, 1983; Shaklee & Mims, 1982; Ward & Jenkins, 1965). Man kann dieses Phänomen auch in der Esoterik-Szene, deren Anteil am Umsatz im Buchhandel inzwischen bei über 30% liegt, beobachten:

Modellbildung durch Statistiken

Ich bezeichne diese Phänomene als „**verfrühte Modellbildung**“, wo in Ermangelung einer soliden Schulausbildung unhaltbare und fehlerhafte Denkmodelle („Memetik“, „Morphogenetische Felder“, „Homöopathie“, „Gott“, „Benjamin Blumchen“, „Pinocchio“...) für „plausibel“ - also in der Wirklichkeit tatsächlich existierend - gehalten werden und unkritisch in die eigenen Denkstrukturen eingebaut werden. Und auf diesem falschen Weltmodell werden dann weitere Ereignisse dann „gemessen“ bzw. mental eingeordnet, wobei diese, je später sie entdeckt werden, umso weniger mehr korrigierbarer sind. Mentale Modelle bauen stets auf mentalen Modellen auf. Oftmals finden sich in unsere Sprache Bezeichner, für welche es kein „Bezeichnetes“ gibt. Falsche Weltmodelle - in frühen Jahren Kindern vermittelt - wirken daher mitunter wie Dynamit. Die Kinder könnten, wie israelische Forscher festgestellt haben, als Heranwachsende im Falle einer empfundenen „Perspektivlosigkeit“ zu tickende Zeitbomben werden, siehe hier: <http://www.little-idiot.de/teambuilding/SchulbuecherPalaestina.pdf>

Illusorische Korrelationen

Welche Erklärung gibt es zu solchen Fehlurteilen hier bei obigem Beispiel der Krankenschwestern? Eine Erklärung von Smedslund ist, dass die Krankenschwestern dazu neigten, bestimmte Informationen unterschiedlich stark zu stark zu gewichten, auch oft „selektive Wahrnehmung“ genannt, eine bereits sokratische Erkenntnis, welche die Rhetorik begründet. Unser Gehirn lernt nicht nur Muster, welche wir wiedererkennen, sondern diese Denklogiken werden noch durch Erziehung emotional gewichtet. Daher handelt sich unser Denken nicht nur entlang der impliziten Logiken bekannter Begriffe, sondern mit Vorliebe an den Denklogiken entlang, welche uns - durch zumeist freundliche und empathische Menschen vermittelt wurden, also emotional „positiv besetzt“ sind. Und in dem gleichen Masse, wie wir etwas lernen, wird unser „Wahrnehmungsfilter“ im Hippocampus programmiert, um uns vor Reizüberflutung zu schützen. Das Phänomen ist bei Kindern zu beobachten, welche - nachdem man ihnen erklärt hat, dass im Wald viele Pilze wachsen, wie sie aussehen, u.s.w. - aus dem Wald nach der Pilzsuche zurückkommend - diese übereinstimmend berichten, dass es im Wald ganz viele Pilze gegeben habe, und zwar „nur“ Pilze. Alle anderen Dinge haben sie kaum wahrgenommen, weil ihre Aufmerksamkeit ganz auf die Wahrnehmung von Pilzen am Boden konzentriert war.

Die Interpretation von Daten wird in Richtung ihrer Vorannahme verzerrt, auch „**Illusorische Korrelation**“ genannt, ein sehr häufig auftretendes Phänomen, siehe *Chapman & Chapman (1969)*.

Und so findet man häufig, dass genau derjenigen Zelle, in der Symptom und Krankheit beide anwesend sind, besonders starke Beachtung zuteil wird. Die Strategie, nur diese Zelle zu beachten oder sie übermäßig zu gewichten, wird tatsächlich oft beobachtet. Eine andere häufige Strategie ist die zu beurteilen, wie viel Prozent derjenigen, die das Symptom haben, auch krank sind. Diese Strategie berücksichtigt die Einträge in der oberen Zeile („anwesend“), vernachlässigt aber völlig die Informationen in der unteren Zeile

(„abwesend“), aus der zu entnehmen ist, wie viele Kranke das Symptom nicht zeigen (vgl. *Arkes & Harkness, 1983; Shaklee & Tucker, 1980*). Diese Strategie ist dieser „Analyse“ ist hier natürlich nicht passend, weil die Tatsache, dass beispielsweise 80 Prozent der Symptomträger krank sind, keineswegs für eine positive Korrelation sprechen muss, auch wenn das Gefühl, also der scheinbar „gesunde“ Menschenverstand dafür spricht. Wenn etwa 100 Prozent aller Personen, die das Symptom nicht zeigen, krank sind, so würde dies offensichtlich ja für eine „negative Korrelation“ sprechen.

Illusion der Kontrolle

Ein weiterer Grund dafür, warum wir Schwierigkeiten mit Scheinkorrelationen und Phantomzusammenhängen haben, liegt auch darin begründet, dass wir zur Illusion neigen, Kontrolle über unsere Umgebung auszuüben, selbst wenn dies gar nicht der Fall ist. *Alloy und Abramson (1979)* haben dies in einem Experiment demonstriert, in dem die Probanden hin und wieder auf eine Taste drücken und daraufhin das Verhalten eines Lämpchens beobachten sollten. Es zeigte sich, dass die Versuchsteilnehmer einen Zusammenhang zwischen Tastendruck und Verhalten des Lämpchens sahen, selbst wenn gar keiner vorlag. Sie hatten also die Illusion, Kontrolle auszuüben. Interessanterweise waren lediglich *Depressive* in der Lage, die „objektiven“ Zusammenhänge korrekt zu erkennen, was in dem Untertitel des Artikels **„Sadder but wiser“** treffend zum Ausdruck gebracht wurde.

Kritische Mitmenschen

Offensichtlich sind wir unter bestimmten Bedingungen jedoch sehr wohl in der Lage, Korrelationen korrekt zu erkennen und zu nutzen (vgl. *Wasserman, 1990; Shanks, Holyoak & Medin, 1996*), andernfalls hätte die Menschheit schwerlich überleben können. Insbesondere sind wir hier auf Kommunikation mit unseren „kritischen“ Menschen, „Bedenkenträgern“, ... angewiesen, die mit ihrer mitunter negativen Weltanschauung erheblich dazu beitragen, indem sie diese Fehler entdecken. Dennoch verfügen wir über diese Kompetenz keineswegs stabil in allen Kontexten, weswegen wir uns darauf nicht verlassen können. Gerade im Bereich wissenschaftlicher Theorienbildung neigen wir zu Verzerrungen und Fehlurteilen. In einigen Wissenschaften sind durch den „Positivismus“ (Glaube nur, was Du sehen/nachweisen kannst) und durch das „Ockham'sche Rasiermesser“ (**„Entia non sunt multiplicanda sine necessitate“** - **„Die Dinge sollten nicht ohne Notwendigkeit vermehrt werden“**) sehr viele Modelle eliminiert worden. Oscar Wilde spricht über den Wissenschaftsprozess so: **„Wissenschaft ist die Geschichte toter Religionen!“** Im Bereich der Physik hat dies dazu geführt, dass im Laufe der Jahrhunderte die Vielzahl von mentalen Modelle der Physik inzwischen mittels der Mathematik „harmonisiert“, nach dem „Kohärenzprinzip“ ineinander überführt werden konnten, bzw. widerspruchsfrei koexistieren können, siehe auch Feyerabend - „Wider den Methodenzwang“. Andere Wissenschaften stehen hier oft noch am Anfang, allen voran die Psychologie als überaus junge Wissenschaft.

Das Testen von Hypothesen

Wissenschaftliches Denken beinhaltet zu einem großen Anteil Prozesse des Hypothesentestens (vgl. *Kuhn, 1997*), sprich Modellfindung, siehe auch *„General System Theory“, Ludwig Von Bartalanffy*. Eine Viele Untersuchungen befassten sich mit der Frage, welche Strategien wir beim Testen von Hypothesen im Alltag einsetzen. Auch hier ist etwas ähnliches zu beobachten, wie bei der *„illusorischen Korrelation“*, nur trifft es hierbei auf die Theorie oder das „mentale Modell“ zu. Ich möchte es daher als *„illusorisches Modell“* oder sogar *„illusorische Modellbildung“* bezeichnen. Die klassische Studie hierzu stammt von *Wason (1960)*. In diesem Experiment wurde Versuchsteilnehmern mitgeteilt, dass die Zahlenreihe 2-4-6 einer bestimmten Regel folge und dass es Aufgabe der Versuchsteilnehmer sei, diese Regel herauszufinden. Nun 2-4-6-8 ist eine naheliegende Annahme, wobei 2-4-6-10 wegen $2+4=6$ und $4+6=10$ eine ebenfalls naheliegende Reihe ist, je nachdem, welches „mentale Modell“ man zugrunde liegt. Wason testete Kandidaten auf die Fähigkeit zur „korrekten Modellbildung“.

Philosophie und mentale Modelle

„Ausnahmen bestätigen die Regel!“ - Volksmund tut Unsinn kund! Ausnahmen widerlegen die Regel! Dennoch sorgt ein Mechanismus, der unter „selektiver Wahrnehmung“ bekannt ist, dafür, dass, obwohl sich diese Regel schon lange nicht mehr in der Wirklichkeit beobachten lässt, also kein Zusammenhang mehr nachweisbar ist, dass das mentale Modell aufrecht erhalten wird. Oftmals artet dies in einen „Generationenkonflikt“ aus, wo der jungen Generation von der alten ein „kollektiver Irrtum“ unterstellt wird. Als Kenner der Wissenschaftsgeschichte, siehe auch Feyerabend, „Wider den Methodenzwang“, stellt man leicht fest, dass bestimmte mentale Modelle z.B. das religiöse Weltbild die zeitgenössischen Weltbilder prägten. Angefangen mit dem geozentrischen Weltbild über das heliozentrische Weltbild bis heute zum „egozentrischen“ ;-). Durchforscht man parallel hierzu die Geschichte der Philosophie, so kann man feststellen, dass hier eine ganze Reihe von Weltbildern und damit verbundene Argumentationsweisen parallel sozusagen in Mode waren:

1. Eine Welt/eine Logik (Tarski, Scholz)
2. Eine Welt/viele Logiken (Grosseteste, Wilson)
3. Viele Welten/eine Logik (Leibniz, Kripke)
4. Viele Welten/viele Logiken (Günther, Derrida).
5. Viele vernetzte Welten und vernetzte Logiken, Auflösung des Gegensatzes von Subjekt/Objekt: Deleuze, Luhmann, Willke, Kiss, Förster, Glasersfeld, Eco, Foucault, Gerhard Roth, Kuhn, Lakatos, Maturana, Vester ...
6. Keine Welt/viele Logiken (Thomas Nagel: „Bedeutsamerweise sind wir unserem Wesen nach auch Betrachter der Welt aus einer Perspektive, die nicht in ihr liegt – aus einer Perspektive von nirgends her in der Welt – und es ist diese Fähigkeit, die es uns ermöglicht, allgemeine und besondere Werturteile zu fällen“)

Auffällig hier der am. Philosoph Thomas Nagel welcher den Gedanken der Kybernetiker aufgreift und konsequent zuende denkt. Die Kybernetiker (siehe Heinz von Förster) haben festgestellt, das unser Gehirn sich stets seine Wirklichkeit stets nur errechnet. Hören wir ein „Miau!“, so erscheint vor unserem geistigen Auge das Bild von Katze. Und dass, obwohl nur Schallwellen an unser Ohr gedrungen sind. Auch die Tatsache, dass wir träumend einen „Bildgenerator“ oder sogar „Moviegenerator“ in uns tragen, sowie das Experiment „Der blinde Fleck“ machen klar, dass unsere Weltanschauung eigentlich nicht von dieser Welt ist. Es fängt schon bei der Betrachtung eines Gegenstandes an, dessen Rückseiten ja verdeckt sind. Der berühmte Philosoph Hegel (aber auch Feuerbach), welcher massgeblich unsere moderneren Weltbilder, z.B. den des „radikalen Konstruktivismus“ (Ernst von Glasersfeld) mitprägte (zurückgeführt auf auch Plato mit seinem Höhlengleichnis) bezeichnete dies als „Abschattung“ (von Denklogiken, Perspektiven, sichtbaren Details).

Modellbildungsfähigkeiten in Führungsjobs

Für Personaler, welche in Assessments über die Einstellung von Bewerbern auf Führungspositionen entscheiden, ist hier diese Frage von entscheidender Bedeutung: Hinterfragen Kandidaten genügend, sind sie kritisch genug? (Siehe Popper mit „Falsifikation“). Es liegt entscheidend an der Erziehung, ob Kinder den Umgang mit Kritik gelernt haben, bzw. ob sie selber ihre „Fehler“ überhaupt bemerken und korrigieren wollen, also Frustrationstoleranz besitzen. Interessant ist auch, dass von ihrer natürlichen Veranlagung depressive Menschen, welche mangels „spielerischer Kreativität“ viele Dinge viel kritischer hinterfragen und somit häufig „geistige Korrekture“ einer Gesellschaft sind (Nietzsche, Tucholski) - „Bedenkenträger“ auch genannt – in einer Unternehmenskultur eine bislang oft unterschätzte, sehr wichtige Rolle als „geistige Korrekture“ spielen. Auch werden sie ungerne in Teams integriert, sind unangenehm hartnäckig, hinterfragen alles, nörgeln herum, „stören“ die „produktive“ Harmonie in einem Kollektiv. Da sie darüber hinaus auch noch öfter krank sind, werden sie von Personalern und ihren „ausgereiften“, „statistischen Verfahren“ in grossen Unternehmen ausfindig gemacht und systematisch herausgekegelt, schliesslich geht es ja um die „Verbesserung des Krankenstandes“. Verbeibend dann noch in der Firma: „Hochmotivierte“, „empathische“ Ja-Sager, Mimosen, leicht zu führen, willig, stets nach Links und Rechts schauend, sich an anderen stets unsicher orientierend. Nur – diese steigern sich gerade in Team – Entscheidungsfindungsprozessen gerne in „kollektive Irrtümer“ hinein.

Argumentationsstrategien

Der berühmte Philosoph Derrida, in Frankreich und U.S.A. - bei den dortigen Juristen z.B. hoch gefeiert - bezeichnete die Marx'sche Dialektik als Seuche, den Versuch, den einen Ansatz durch den anderen überwinden zu lassen, siehe: www.little-idiot.de/teambuilding/DualistischAmbivalentDialektisch.pdf

„Die Behauptung von Gegensätzen dient dazu, uns darauf festzulegen, entweder das eine oder das andere zu sein. Weiss oder Schwarz, Mann oder Frau, eindeutig männlich oder bzw. weiblich oder „gestört“, heterosexuell oder homosexuell, nicht behindert oder behindert ... sind Einteilungen die asymmetrische Machtverhältnisse verkörpern, da jeweils eine Kategorie des Gegensatzpaars die andere dominiert.“ (Jaques Derrida).

Das Grundproblem der demokratischen Entscheidungsfindung: Die einen liegen falsch, die anderen liegen falsch, und sie bilden einen Kompromiss – sie einigen sich auf der Mitte! „Diskursfähigkeit“ ist eine Grundvoraussetzung, an Entscheidungsfindungsprozessen in Teams teilnehmen zu können. Grundlage hierbei ist, Denklogiken und besonders Hintergründe bzgl. der Darstellung von Sachverhalten auszutauschen. „Eine Landkarte ist nicht das Gebiet, das sie repräsentiert, aber wenn sie korrekt ist, ist sie in ihrer Struktur der Struktur des Gebietes gleich (oder ähnlich), worin ihre Brauchbarkeit begründet ist“ - „The map is not the territory!“ (Korzybski). Viele der sachlichen Denklogiken haben eine emotionale Anknüpfung erfahren, was ein Modell der „emotionalen Landkarten“, einem Netzwerk von Emotionslogiken begründet, welche oft nicht mit dem Netzwerk von Sachlogiken übereinstimmt. Wie jemand argumentiert, ist durchaus mit o.a. „weltanschaulichen Modellen“ begründbar. Die grössten

„mentalen Widerstände“ mitsamt dogmatischem Verhalten - hinten herum, also nicht geradlinig nach vorne heraus argumentierend - finden sich bei Vertretern der schwarz/weiss - Denke und den „Himmel/Hölle - gut/böse“ - Vertretern, welche es gut verstehen, in „modernen Unternehmen“ ihre tiefstehenden Denkstrukturen (darunter auch Hierarchiedenken, Zuständigkeitsdenken, Traditionsdenken, Pöstchendenken) zu verbergen. Diese Widerstände zeigen sich, wenn „prozessuales Denken“ in Unternehmen, das Denken in Wertschöpfungsketten (Lean, Kaizen) eingeführt werden soll und auch bei eng vernetzt arbeitenden Teams, z.B. in Programmiererteams beim Übergang von „imperativem/objektorientiertem Programmieren“ zu „funktionalem Programmieren“.

Dominante Strategien in Kollektiven, Menschenführung

In der chinesischen Kultur wird von „sich selbst verzeihen“ gesprochen, also von der Fähigkeit, auch als Erwachsener sich seiner Fehler anzunehmen und seine mentalen Modelle korrigieren zu können. Hierbei werden „strategische Pläne“, also im Grunde „Methodenwissen“ über langfristig bewährte Methoden, sog. „dominante Strategien“ in der Gesellschaft, durch flexible und einfache Leitprinzipien („Simple Guiding Principles“) ersetzt, denen vorrangig eines der o.a. „mentalen Bilder“ zugrunde gelegt wird. Dies begründet die Beliebtheit von bestimmten Weltanschauungen, „Redewendungen“, „Sprüchen“, „Aphorismen“, aber auch Gedichten, Reimen, Literatur, wie z.B. Goethes Faust, welcher in hochkomprimierter Form eine Unmenge von „Lebensweisheiten“ vermittelt. Bestimmte „Leitprinzipien“ sind geeignet, in Form von „Gedankenkeimen“ vermittelt, langfristig die Denkstrukturen eines Menschen zu verändern. Dreh - und Angelpunkt ist es, logische Widersprüche in den Denkstrukturen ausfindig zu machen, und das Gegenüber damit in der Praxis zu konfrontieren, was Stress, also „kognitive Dissonanz“ (siehe Leon Festinger) erst einmal auslöst. Durch geschicktes „Umfahren“ dieser neuralgischen Punkte mittels Sprachlogiken entlang der „freien Strecken“ der emotionalen Landkarten lassen sich dann doch neue mentale Modell implementieren. Hierin liegt die hohe Kunst der Menschenführung begründet.

Ausweichlogiker und Emotionslogiker

Kuhn (1997) ließ Probanden herausfinden, welche Faktoren (Humor, Haarlänge, Tag der Sendung, Musik) die Beliebtheit einer Fernsehsendung beeinflussten. Die meisten Versuchsteilnehmer hielten den Humor für den entscheidenden Faktor. Die erfundenen Daten, die ihnen gezeigt wurden, suggerierten, dass kein Zusammenhang existierte. Sie hielten dennoch an ihrer Hypothese fest und gaben komplizierte Gründe an, warum wohl man keine klare Evidenz für ihre Hypothese sehen konnte. Sie begannen, zu „Ausweichlogikern“ zu werden: „Kann nicht kommen, Lüge kommt später!“.

Solche Kandidaten, emotional zumeist recht weich („**Emotionslogiker**“), sehr umgänglich, empathisch, können daher nirgendwo in Führungsjobs gebraucht werden, wo es darum geht, „hart“ an den tatsächlichen Fakten der Wirklichkeit zu bleiben. Sie besitzen keine ausreichenden Methoden des Abgleiches von ihrer gedanklichen Vorstellung, also ihrer „Realität“ mit der tatsächlichen Wirklichkeit! Wenn untere und mittlere Führungskadres für das Management wichtige, entscheidende Tatsachen verschweigen oder sogar beschönigen, muss die Modellbildung über Tatsachen der Wirklichkeit durch die oberen Führungsetagen fehlschlagen.

Dieser Artikel ist unter

<http://www.little-idiot.de/teambuilding/StatistikenModellbildungMethodenManagement.pdf> nachlesbar.

Mit freundlichen Grüßen, Guido Stepken

Literatur

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Alloy, L. B. & Abramson, L. J. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 441-485.
- Appleton, D. R., French, J. M. & Vanderpump, M. P. (1996). Ignoring a covariate: An example of Simpson's paradox. *American Statistician*, 50, 340-341.
- Arkes, H. R. & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, 112, 117-135.
- Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge: Cambridge University Press.
- Bartalanffy, Ludwig von. *General System Theory*
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Bordens, K. S. & Abbott, B. B. (1999). *Research design and methods*(4th ed.). London: Mayfield.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Berlin: Springer.
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research in teaching. In N. L. Gage (Ed.), *Handbook of research in teaching* (pp. 171-246). Chicago: Rand McNally.
- Nachdruck als Monographie: (1966). *Experimental and quasi-experimental designs for research in teaching*. Chicago: Rand McNally.
- Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid diagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405.
- Dawes, R. M. (1979). The robust beauty of improper linear models. *American Psychologist*, 34, 571-582.
- Dawes, R. M. (2001). *Everyday irrationality*. Boulder: Westview Press.
- Dupré, J. (1993). *The disorder of things. Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.
- Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.
- Fischhoff, B. & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90, 239-260.
- Feyerabend, Paul. *Wider den Methodenzwang*
- Frick, R. W. (1998). Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments & Computers*, 30, 527-535.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 1-13.
- Gigerenzer, G., Todd, P. M. & ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford: University Press.
- Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In G. Lüer (Hrsg.), *Allgemeine Experimentelle Psychologie* (S. 43-264). Stuttgart: Fischer.
- Hager, W. (1992). *Jenseits von Experiment und Quasi-Experiment. Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen*. Göttingen: Hogrefe.
- Hell, W., Fiedler, K. & Gigerenzer, G. (1993). *Kognitive Täuschungen*. Heidelberg: Spektrum Akademischer Verlag.
- Holyoak, K. J. & Thagard, P. (1995). *Mental leaps*. Cambridge, MA: MIT Press.
- Huber, O. (1995). *Das psychologische Experiment* (2. Aufl.). Bern: Huber.
- D. Kahneman, P. Slovic, & A. Tversky (Eds.)(1982), *Judgment under uncertainty: Heuristics and biases*. Cambridge, MA: Cambridge University Press.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Monterey, CA: Brooks/Cole.
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.
- Kuhn, D. (1997). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), *Modes of thought: Explorations in culture and cognition* (pp. 261-281). New York: Cambridge University Press.
- Loehlin, J.C. (1998). *Latent variable models. An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Medin, D. L., Ross, B. H. & Markman, A. B. (2001). *Cognitive psychology* (3rd ed.). Fort Worth : Harcourt College Publishers.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379-387.
- Mynatt, C. R., Doherty, M. E. & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85-95.
- O'Reilly, R. C. & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Popper, K. R. (1994). *Logik der Forschung* (10. Aufl.). Tübingen: Mohr.

- Rosnow, R. L. & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin*, *105*, 143-146.
- Shaklee, H. & Mims, M. (1982). Sources of error in judging event covariation: Effects of memory demands. *Journal of Experimental Psychology: Human Learning and Memory*, *8*, 208-224.
- Shaklee, H. & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, *8*, 459-467.
- Shanks, D. R., Holyoak, K. J. & Medin, D. L. (1996), *The psychology of learning and motivation*, Vol. 34: *Causal learning*. San Diego: Academic Press.
- Shaughnessy, J.J. & Zechmeister, E.B. (1997). *Research methods in psychology*. Boston: McGraw Hill.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, *13*, 238-241.
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, *4*, 165-173.
- Snyder, M. & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, *14*, 148-162.
- Stangl, Werner (1989). *Das Problem des Induktivismus*
- Sternberg, R. J. & Ben-Zeev, T. (2001). *Complex cognition. The psychology of human thought*. New York: Oxford University Press.
- Thagard, P. (2000). *How scientists explain disease*. Princeton: Princeton University Press.
- Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation*, Vol. 34: *Causal learning* (pp. 47-88). San Diego: Academic Press.
- Waldmann, M. R. (1997). Wissen und Lernen. *Psychologische Rundschau*, *48*, 84-100.
- Waldmann, M. R. & Hagmayer, Y. (in press). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*.
- Ward, W. C. & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231-241.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.
- Wasserman, E. A. (1990). Detecting response-outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 26 (pp. 27-82). New York: Academic Press.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik. Ein Lehrbuch zur Psychologischen Methodenlehre*. Göttingen: Hogrefe.
- Zaltman, Gerald. *How customers think!*