

Desorientierung durch falsche Maße¹

Peter Jaenecke

Rationale Entscheidungen erfordern gesicherte Kenntnisse über den Istzustand des jeweiligen Tätigkeitsfeldes; sie werden durch Messungen gewonnen. Messen beruht auf allgemeinen und für alle Fachgebiete verbindlichen Messmethoden. Eine der wichtigsten Methode, die Verwendung eines Maßes, ist der Ausgangspunkt dieser Untersuchung; sie wird zunächst an einem einfachen physikalischen Fall erläutert. Anschließend wird anhand von einfachen alltäglichen Beispielen gezeigt, dass sie außerhalb der Naturwissenschaften ziemlich unbekannt zu sein scheint, vermutlich deshalb, weil man die Maßproblematik fälschlich für eine rein quantitative Angelegenheit hält. Doch sie tritt auch bei qualitativen Beziehungen auf, denn die Forderung, korrekte Maße festzustellen, bedeutet nichts anderes als kausale Zusammenhänge aufzudecken. Aus diesem Grund birgt solch ein Korrektheitsnachweis ein verkanntes und daher weitgehend ungenutztes schöpferisches Potenzial. Viele Messungen/ Datenerhebungen beruhen zunehmend auf falschen Maßen und/oder auf methodischen Defiziten im Umgang mit Ihnen. Die so gewonnenen Messergebnisse liefern ein verfälschtes Bild von der Wirklichkeit, dennoch dienen sie allerorten als Entscheidungsgrundlage. Daraus entstehen zwangsläufig Wirklichkeitsverlust und Desorientierung, die zu langfristigen Fehlentwicklungen in Wissenschaft und Gesellschaft können. Im Anhang werden hierfür aktuelle Beispiele aufgelistet, welche, wie das Hochschul-Ranking, gegenwärtig Gegenstand kontroverser Debatten sind.

1 Einleitung

Früher beruhten Entscheidungen vorwiegend auf persönlichen Lebenserfahrungen; in unserer hochtechnisierten, komplexen Welt vermag sich jedoch ein Einzelner kaum noch aus eigener Anschauung ein Urteil zu bilden. Unser Handeln gründet sich heute hochgradig auf sprachlich übermitteltes theoretisches Fremdwissen. Doch diese theoretische Basis und die aus ihr abgeleiteten Vorstellungen über das, was wünschenswert und machbar sein sollte, ermöglichen allein

noch kein vernunftgeleitetes Handeln; es muß noch etwas anderes hinzukommen: Die Kenntnis des Istzustandes. So lässt sich mit der Himmelsmechanik, für sich genommen, wenig anfangen; um die Stellung eines Planeten für alle zukünftigen und vergangenen Zeitpunkte berechnen zu können, muß zunächst der Zustand des Planeten zu einem ganz bestimmten, beliebig wählbaren Zeitpunkt bekannt sein. Durch diese Daten wird die Verbindung zwischen der allgemeinen Theorie (der Himmelsmechanik) und dem speziellen Fall (der Bahn des fraglichen Planeten) hergestellt. Solch eine das Verständnis der Zusammenhänge voraussetzende Zuordnung ist auch dann erforderlich, wenn die Hypothesen vage und wenig gesichert sind. Entscheidend ist hier nämlich zunächst nicht deren Güte, sondern die Tatsache, dass es sich um allgemeine Aussagen handelt, die sich erst nutzen lassen, wenn sie mit speziellen Daten verknüpft werden:

Wer rational urteilen und handeln will, benötigt neben den allgemeinen theoretischen Kenntnissen auch spezielle Angaben über den Istzustand seines Wirkungsfeldes.

Bei der Anwendung von theoretischem Wissen wird daher die Frage, wie solch ein Istzustand zu bestimmen sei, zu einem zentralen Problem.

Für Naturwissenschaftler ist dies eine triviale Feststellung; andernorts fehlt hierfür häufig jedes Problembewusstsein, vielleicht deshalb, weil man übersieht, dass im Wissen unausgesprochene Hypothesen verankert sind, deren Gebrauch man sich nicht bewusst ist. In vielen Fällen geht eine Entscheidung nur von subjektiven Wahrnehmungen oder bloßen Vermutungen aus. Man spricht zwar mitunter von Bestandsaufnahme und Istanalyse, doch beschränken sich solche Aktivitäten oft auf zweifelhafte statistische Erhebungen, auf Eignungstests, Umfragen, Bestimmung von volkswirtschaftlichen Indikatoren usw., die, wenn überhaupt, meist nur einen ganz bestimmten besonders augenfälligen Aspekt des Istzustandes erfassen.

¹ Erweiterte Fassung eines Vortrags, gehalten auf der Tagung über Begriffliche Wissensverarbeitung, Technische Hochschule Darmstadt, 23.- 26. Februar 1994.

Rechnete man bei astronomischen Berechnungen ebenso oberflächlich etwa mit bloß mutmaßlichen Anfangswerten, kämen falsche Bahnen heraus, obwohl den Berechnungen eine gesicherte Theorie zugrundeliegt. Um wieviel fragwürdiger müssen dann aber erst die Ergebnisse sein, die mit unzuverlässigen Daten und unsicheren Hypothesen gewonnen wurden! Dennoch werden anhand solcher Ergebnisse Entscheidungen getroffen, vor allem im gesellschafts- und wirtschaftspolitischen Bereich, wo Fehlentscheidungen für viele Menschen schwerwiegende und, in zunehmendem Maße, irreversible Folgen haben können. Eingedenk dessen muss man die häufig zu beobachtende unkritische Beschaffungspraxis von Erfahrungsdaten als grob fahrlässig einstufen.

Zur Entschuldigung muß allerdings gesagt werden, dass die Problematik der Datengewinnung bereits von der hierfür zuständigen, von der Tradition abhängigen, praxisfernen Wissenschaftstheorie sträflich vernachlässigt wird. Ein Wissenschaftler, der seine Arbeit ernst zu nehmen versuchte, würde bei ihr vergeblich um Rat nachfragen. Vor allem scheint man bisher nicht erkannt zu haben, dass jede Form von systematischer Datengewinnung eine Messung ist:

Zuverlässige Erfahrungsdaten, gleichgültig ob physikalische oder nicht-physikalische, können nur durch Messungen gewonnen werden.

Dies ist der Ausgangspunkt für die folgenden Untersuchungen. Wir beginnen mit der Erläuterung von Maß und Maßfunktion, zeigen, dass die Verwendung eines Maßes eine allgemeine Messmethode ist und beschreiben anhand von alltäglichen, oft umgangssprachlich versteckten Beispielen die desorientierenden Wirkungen, die aus dem Gebrauch von falschen Maßen und aus methodischen Defiziten im Umgang mit ihnen hervorgehen. Zum Schluss weisen wir auf die tieferliegenden Zusammenhänge der Maßproblematik hin und kommen zu dem Ergebnis, dass qualitative und quantitative Beziehungen eine Einheit bilden. Resümierend müssen wir feststellen, dass viele Fehlentwicklungen unserer Tage auf die Verwendung von falschen Maßen zurückgehen und eine Scheinobjektivität vorspiegeln. Gerade weil die Maßproblematik gegenwärtig so sehr vernachlässigt wird, bietet die Auseinandersetzung mit ihr in den einzelnen Wissenschaften eine vielversprechende Anregung für neue, schöpferische Ideen.

2 Maß und Maßfunktion

Um zu vermeiden, dass ein Schiff in einem unbekanntem Gewässer auf Grund läuft, sendet man ein akustisches Signal in Richtung Meeresboden, misst die Zeitdauer, die zwischen dem Senden und dem Empfang des Echos vom Meeresboden vergeht und bestimmt daraus die Wassertiefe. Dieses einfache, aber nichttriviale Messverfahren gewährt einen Einblick in ein grundlegendes Messprinzip:

Die gesuchte Größe, die Wassertiefe, ist eine Länge; gemessen wird jedoch eine Zeitdauer. Wir bezeichnen bei dieser Messung die Zeitdauer als (empirisches) *Maß* für die Wassertiefe. Länge und Zeitdauer kennzeichnen zwei völlig verschiedene physikalische Merkmale, dennoch kann das eine das Maß des anderen sein. Dies ist jedoch nur dann der Fall, wenn zwischen beiden Größen ein naturgesetzlicher Zusammenhang besteht; seine mathematische Darstellung bezeichnen wir als *Maßfunktion*.² Bei der Echolotung ist es das bekannte Gesetz

$$s = \frac{1}{2} v t,$$

also 'Weg ist gleich Schallgeschwindigkeit (im Wasser) mal Zeitdauer', versehen noch mit dem Faktor $\frac{1}{2}$, da der Schall die fragliche Strecke zweimal zurücklegt; t sind Messwerte vom Maß 'Zeitdauer', s die berechneten Messwerte von der gesuchten Größe 'Wassertiefe'.

Ein Maß ist ein Merkmal, für das es bereits ein Messverfahren gibt; es ist also auch seinerseits nur über ein Maß zugänglich. So ist die Zeitdauer kein beobachtbares Merkmal und wird daher z.B. über das Zifferblatt auf eine Winkel- bzw. Längenmessung zurückgeführt. Ein Maß für eine zu messende Größe finden und die zugehörige Maßfunktion bestimmen, ist oft ein sehr mühevoller Unterfangen.³ Lässt es sich nicht durch eine direkte Messung umgehen?

² Da eine Messung immer auf einem Gesetz beruht, kann es keine Asymmetrie zwischen Verifikation und Falsifikation geben. POPPER, auf dieser Asymmetrie aufbauend, hat daher das Induktionsproblem nicht gelöst.

³ Viele Vorhaben bleiben dadurch auf der Strecke. Dem Leistungsprinzip wieder Geltung zu verschaffen, scheitert z.B. vor allem daran, dass man nicht weiß, wie man Leistung bemessen soll.

Unter 'direkt Messen' könnte man 'Messen ohne einen Vergleich auszuführen' verstehen. Doch wollte man nach dieser Art etwa die Länge eines Tisches bestimmen, so hieße das z.B. Kerben in den Rand des Tisches ritzen und die Zahl der Kerben zählen. Man könnte dann sagen, der Tisch sei soundsoviele Kerben lang, doch solch eine singulären Aussage wäre völlig wertlos; ein Tischler, beauftragt, einen gleich langen weiteren Tisch zu bauen, könnte mit ihr nichts anfangen. Messen hat nur Sinn, wenn die Gegenstände an einem zuvor vereinbarten Maß gemessen werden, denn dann sind auch die Messwerte miteinander vergleichbar. Dem begrifflichen Vergleich auf der Zahlenebene muß also immer ein empirischer Vergleich auf der Handlungsebene vorausgehen, d.h. ein Zollstock muß an die Tischkante angelegt, die Kantenlänge auf den Zollstock übertragen und dort abgelesen werden:

Die Verwendung eines Maßes ist eine allgemeine, für alle Arten von Messungen gültige Messmethode.⁴

Eine direkte Messung, verstanden als eine Messung ohne die Vermittlung durch ein Maß, gibt es nicht. Jede Messung setzt mit dem Maß auch eine Maßfunktion voraus; dies wird häufig übersehen, weil die Maßfunktion als Skala am Messgerät erscheint. So ist die Temperaturmessung eigentlich eine Längenmessung, durch die Skala am Thermometer lässt sich aber unmittelbar die Temperatur ablesen. Eine Messung besteht daher immer aus zwei Schritten: aus einem experimentellen Teil, bei dem die Messung vorgenommen und einem rechnerischen, bei dem der Messwert berechnet wird. Bei der Echolotung wird t gemessen und daraus s berechnet; v muß als bekannt vorausgesetzt werden.

Wie bei der Echolotung unterscheiden sich bei vielen Messungen die Merkmale, auf denen Maß und zu messende Größe beruhen. 'Direkt' könnte auch heißen 'Maß und zu messende Größe beziehen sich auf das gleiche Merkmal'. Ein Zollstock z.B. repräsentiert das physikalische Merkmal 'Länge', und mit diesem Messgerät werden selbst wieder Längenbestimmungen ausgeführt. Solche Messverfahren, bei denen die Maßfunktion stets gleich der Identität ist, sind möglich, wie

⁴ Neben der Verwendung eines Maßes gibt es noch weitere Prinzipien, z.B. solche, die sich auf die Messfehlerproblematik beziehen; sie werden ausführlicher behandelt in JAENECKE (2014): *Einführung in die Messtheorie*. Wir beschränken uns hier auf die Maßproblematik. Auch die anderen Prinzipien werden wenig beachtet und führen zu zahlreichen Irrtümern.

das soeben erwähnte Beispiel zeigt; es muss sogar mindestens ein Verfahren dieses Typs geben, um dem unendlichen Regreß zu entgehen, der darin bestünde, dass ein Maß, um eine messbare Größe zu sein, selbst wiederum eines Maßes bedarf.⁵

Die Frage ist nur: Gibt es stets für jedes Messvorhaben ein Maß, das sich auf das gleiche Merkmal wie die zu messende Größe bezieht? Die Wassertiefe lässt sich in der Tat in diesem Sinn direkt mit einem Senkblei ermitteln, dennoch zieht man heute die Echolotung vor, weil sie viel leichter zu handhaben ist und erheblich genauere Werte liefert. Hier greift man auf ein anderes Merkmal zurück, weil so die fragliche Größe besser zugänglich ist; im Prinzip aber wäre solch ein Wechsel nicht erforderlich. Dies gilt aber nicht für alle Messvorhaben. Messungen werden durchgeführt, um uns Kenntnisse der uns umgebenden Welt zu verschaffen. Diese Welt ist uns aber nur insoweit zugänglich, als uns unsere Sinnesorgane diesen Zugang verschaffen. Alles was außerhalb ihres Empfangsbereiches liegt, ist uns nicht direkt zugänglich. Wir werden niemals eine Temperatur wahrnehmen⁶ oder die Lebenshaltungskosten wachsen sehen. Wenn wir uns aber trotzdem Zugang zu diesen Bereichen verschaffen wollen, benötigen wir ein entsprechendes Maß: für die Temperatur z.B. das temperaturabhängige Volumen einer Flüssigkeit, für die Lebenshaltungskosten einen Warenkorb. Da gerade die nicht-physikalischen Größen nicht direkt beobachtet werden können, sind sie uns, wenn überhaupt, *nur* über ein Maß zugänglich. Doch eine Größe z ist nur dann ein Maß für eine gesuchte Größe x , wenn es für sie eine Maßfunktion M der allgemeinen Form

$$x = M(z)$$

gibt, welche die beiden Anforderungen erfüllt:

⁵ Welches Merkmal als Grundmaß verwendet werden sollte, hängt von den Wahrnehmungsfähigkeiten des Messenden ab. Beim Menschen ist der Sehsinn am ausgeprägtesten, deshalb besitzt für ihn die Längenmessung eine besondere Bedeutung; in technischen Systemen dominiert dagegen die Messung einer elektrischen Spannung.

⁶ Der Mensch besitzt kein Temperaturempfinden; was er wahrnimmt ist ein Energiefluß: Ein Gegenstand fasst sich kalt an, wenn er dem Körper Wärmemenge entzieht, warm, wenn er dem Körper Wärmemenge zuführt.

1. M muss einen Gesetzeszusammenhang zwischen x und z beschreiben.
2. M muss eine streng monotone Funktion sein.

Dann ist nämlich auch die Inverse M^{-1} streng monoton und erfüllt folgende Symmetriebedingung: wenn z ein eindeutiges Maß für x ist, so ist umgekehrt auch x ein eindeutiges Maß für z gemäß

$$z = M^{-1}(x).$$

Im Folgenden beschäftigen wir uns mit den Problemen, die sich aus der Missachtung dieser Anforderungen ergeben; dabei richten wir unser Augenmerk insbesondere auf Güte- bzw. Leistungsbemessungen.

3 Interpretation von Messergebnissen

Messwerte sind zunächst nichts weiter als Zahlen. Wenn man sie nutzbringend anwenden will, muss man sie daher erst in geeigneter Weise interpretieren. Die Interpretation von Messwerten ist daher ein weiterer Problemschwerpunkt bei der Ermittlung des Istzustandes. In vielen Fällen werden zunächst irgendwelche Daten erfasst, man weiß aber nicht, auf welche Größe sie sich beziehen; d.h. das Maß ist bekannt, es liegen auch Messwerte vor, aber es herrscht Ungewißheit darüber, für welche Größe es ein Maß ist. Der Anteil der Nichtwähler z.B. lässt sich problemlos bestimmen, doch über seine Bedeutung gehen die Ansichten weit auseinander; am häufigsten wird Politikverdrossenheit vermutet, doch er könnte u.a. auch durch die Unattraktivität der Parteien bedingt sein. Die Kontroversen und häufig auch irigen Ansichten über das fragliche Phänomen offenbaren den theoretischen Charakter des Problems. Es wird um die Ursachen gestritten, wobei zahlreiche sich widersprechende Hypothesen ins Spiel kommen; das zugrundeliegende messtheoretische Problem bleibt jedoch unerkannt, ebenso, dass es Maß und Maßfunktion sind, die den sinnstiftenden Zusammenhang vermitteln. 'Interpretieren' ist hier eine irreführende Bezeichnung; sie weckt den Eindruck, es gebe einen Interpretationsspielraum, der sich argumentativ eingrenzen ließe. In Wirklichkeit handelt es sich aber bei der Messwert-Interpretation formal um eine Schlussfolgerung der Form:

Allgemeines Wissen/Theorie
Messergebnis(se)

Konklusion

Oberhalb des Schluss-Striches stehen die als wahr angenommenen Prämissen oder Voraussetzungen, unterhalb von ihm steht die Konklusion. Eine gültige Schlussfolgerung erfordert die Wahrheit aller Prämissen; dann ist auch die Konklusion wahr.

Beispiel:

- P₁ Wenn Objekt A höher ist als Objekt B, kann man das Objekt B in das Objekt A hineinstellen.
P₂ Ein Objekt ändert seine Höhe nicht, wenn es von einem Ort zu einem anderen transportiert wird.
P₃ Das Zimmer A ist höher als die Schrankwand B.

K Also passt die Schrankwand B in das Zimmer A.

Die ersten beiden Prämissen werden als trivial empfunden und daher selten eigens erwähnt; Man setzt sie stillschweigend als gültig voraus. Stillschweigende Voraussetzungen, aber auch plausibel erscheinende Hypothesen und/oder ideologisch bedingte Aussagen können jedoch schnell zu einer voreiligen Fehlinterpretation führen; sie hat mit

Ausgewähltes Wissen
stillschweigende unzutreffende Vermutungen
plausible erscheinende Hypothesen
ideologisch bedingte Annahmen
Messergebnis(se) (ohne Fehlerangaben)

Fehlinterpretation

ebenfalls die Form einer Schlussfolgerung, allerdings mit mangelbehafteten Prämissen: die Mängel wurden im obigen Schema rot markiert. Um sich vor Fehlinterpretationen und letztlich vor Fehlentscheidungen mit weitreichenden Folgen zu schützen, ist es daher dringend erforder-

lich – entweder gestützt auf eine anerkannte Theorie oder über weitere empirische Untersuchungen –

- (1) sich über die verwendeten Prämissen Klarheit zu verschaffen und
- (2) ihre Wahrheit zu sichern.

Das gelingt aber nur, wenn man weiß, wie die Prämissen zustandegekommen sind. Das dazu erforderliche Wissen steckt aber in der Maßfunktion. Sie hat daher eine Doppelfunktion: Zum einen muss sie korrekte Werte garantieren, zum anderen muss sie für eine Interpretation hinreichend informationsreich sein.

Bei Nachweisexperimenten liegt die Konklusion bereits fest; sie lautet: 'also existiert ...' Der experimentelle Aufbau wird daher so ausgerichtet, dass sich bei entsprechenden Messergebnissen automatisch die Konklusion ergibt. Hier besteht die Gefahr, dass diese Vorgehensweise die Form einer selbsterfüllenden Prophezeiung annehmen kann.

Die Interpretation von Messwerten ist gewissermaßen die umgangssprachliche Erläuterung der Maßfunktion. Umgekehrt bedeutet dies: Auch wenn man sich auf solch eine Interpretation einlässt, postuliert man stillschweigend eine ganz bestimmte Maßbeziehung. Entscheidend ist nämlich nicht, ob man sich einer solchen Funktion bewusst ist oder nicht, sondern welches Gesetz gelten muß, damit die betreffende Interpretation einen Sinn bekommt. Wenn man z.B. behauptet, Wahlenthaltung zeuge von Politikverdrossenheit, so hat man automatisch einen Gesetzeszusammenhang zwischen Wahlenthaltung und Politikverdrossenheit postuliert, andernfalls wäre ja die Aussage sinnlos. Vom Autor unbemerkt, stehen solche impliziten Hypothesen oft mit seinen explizit geäußerten im Widerspruch, insbesondere dann, wenn die Messergebnisse als Beweis für oder gegen eine These ins Feld geführt werden.

4 Desorientierung beim Gebrauch von falschen Maßen

Dass irgendein Zusammenhang zwischen den am Messvorgang beteiligten Größen bestehen muss und somit auf die erste Anforderung zumindestens im Kern nicht verzichtet werden kann, darüber herrscht weitgehende Übereinstimmung. In der Praxis behilft man sich damit, umgangssprachlich formulierte gesetznährliche Aussagen zu treffen, die einen kausalen Zusammenhang suggerieren. Ob solch ein kausaler Zusammenhang überhaupt besteht und wenn ja, unter welchen Bedingungen, wird nicht überprüft. Manche Formulierungen deuten daraufhin, dass sich ihre Urheber selbst nicht so ganz imklaren über die Tragweite ihrer Behauptungen sind und sich mehr auf ihr Gefühl als auf ihren Verstand verlassen. Weitgehend übersehen wird dabei, dass die in der Natur ablaufenden kausalen Vorgänge immer an quantitative Veränderungen gekoppelt sind und dass deren Eigenschaften bekannt sein müssen, wenn man sinnvoll messen will.

Sind nun das Maß (und somit auch die Maßfunktion) unzulässig, so sind es auch alle nachvollgehenden Anwendungen der Messergebnisse. Hält man dennoch an ihnen fest, so ist eine Desorientierung die Folge: man richtet sich in seinen Entscheidungen gewissermaßen nach einem falsch anzeigenden Kompass und läuft dabei zwangsläufig in die Irre, manchmal, ohne es mangels Korrekturmöglichkeiten selbst zu bemerken. Das möchte man natürlich so gut es geht vermeiden, doch eine absolute Sicherheit gibt es nicht: auf Holzwege kann man immer geraten, aber man ihre Zahl eindämmen, und zwar durch Ausprägung eines scharfen Problembewusstseins. Doch damit hapert es bereits gewaltig im Wissenschaftsbetrieb; dieses zeigen die Beispiele aus dem ersten Abschnitt. In den umgangssprachlich formulierten kausalen Zusammenhängen findet man oft Voraussetzungen und Interpretation in unzulässiger Weise miteinander vermengt.

Im zweiten Abschnitt geht es um die Problematik, welche Maße mit sich bringen, die – eventuell gestützt auf eine anerkannte Theorie – durch eine Definition eingeführt wurden. Nach unserem Verständnis bedeutet dies: ein Gesetzeszusammenhang wurde postuliert. Der dritte Abschnitt beschäftigt sich mit einem definierten Maß besonderer Art, das heute große Konjunktur besitzt: es ist der Score. Die zur Erläuterung verwendeten Beispiele orientieren sich hauptsächlich an Güte- bzw. Leistungsmessungen; es sind die häufigsten und im gesellschaftlichen Leben zugleich bedeutendsten nicht-physikalischen Messungen. Immer wenn es gilt, nicht willkürlich, sondern nach objektiven Kriterien den geeignetsten Bewerber für ein Amt, die beste Lösung für

ein gegebenes Problem, das Produkt mit der höchsten Qualität, die Maßnahme mit dem größten Nutzen usw. herauszufinden, so ist in irgendeiner Form eine Messung im Spiel; deren Ergebnisse dienen als Grundlage für viele wichtige Entscheidungen.

Wir beginnen mit einigen Beispielen, die zeigen, wie unbedarft man vor allem in nicht-physikalischen Lehrbüchern durch umgangssprachlich salopp formulierte „Maßaussagen“ mit der Problematik umgeht.

4.1 Methodische Defizite beim unbedarften Umgang mit Maßen

Die nicht-physikalische Literatur ist reich an Beispielen, in denen unbewusst von Maßen Gebrauch gemacht wird. Sie müssen deshalb nicht falsch sein; doch als Leser wünscht man sich zum besseren Verständnis nicht bloß Behauptungen wie *selten sind kreative Persönlichkeiten auch konsistente Persönlichkeiten*,⁷ sondern auch deren Rechtfertigungen. Sie werden oft nicht als notwendig erachtet; sprachliche Formulierungen wie 'das Sein bestimmt das Bewusstsein' täuschen jedoch über den quantitativen Charakter der Aussage und dessen Problematik hinweg; gemeint ist anscheinend etwas viel Weitergehendes, nämlich 'je schlechter die materielle Lage eines Menschen, desto ausgeprägter sein Klassenbewusstsein.' Auch in Definitionen finden sich oft verdeckte quantitative Anspielungen. Die sprachliche Einkleidung verschleiert offenbar die Tragweite des Behaupteten:

*GUILFORD ging davon aus, dass jeder Mensch bestimmte Merkmale aufweise, die sehr eng mit der Kreativität gekoppelt seien. Eine hiervon ist die Sensibilität - und zwar bezogen auf bestimmte Sachen. Beispielsweise ließ er die Testpersonen zu einem vorgelesenen Text beliebig viele Kommentare machen oder Fragen stellen. Dies sagte viel über deren Einfühlungsvermögen aus. Ferner spielte der sogenannte Ideenfluß eine entscheidende Rolle. Mit anderen Worten: wie schnell und wie viele originäre Ideen bei jedem einzelnen abrufbar sind.*⁸

Hier wird zunächst behauptet, Kreativität habe etwas mit Sensibilität zu tun; die Kommentare und Fragen zu vorgelesenen Texten gelten dabei als maßgebend für das Einfühlungsvermögen.

⁷ NEUBEISER (1993): *Die Logik des Genialen*, p. 52. Das Buch hält bei weitem nicht das, was sein Titel verspricht.

⁸ NEUBEISER (1993): *Die Logik des Genialen*, p. 63f.

Kreativität wird aber auch mit dem Ideenfluss in Verbindung gebracht und dieser wird daran gemessen, wie schnell und wie viele Ideen – unabhängig von ihrem Gehalt – abrufbar sind. Mehr als eine oberflächliche Plausibilität können solche Behauptungen nicht beanspruchen.

Methodische Defizite gibt es auch in Bezug auf die Maßfunktion. Selbst wenn zwei Größen etwas miteinander zu tun haben, so ist damit noch nicht gesagt, dass sich auch die eine als Maß für die andere eignet, denn das Gesetz, das die beiden Größen miteinander verknüpft, muß, um als Maßfunktion infrage zu kommen, eine im Messbereich streng monotone Funktion sein, und je nach ihren übrigen Eigenschaften sind bestimmte Vergleiche zwischen den Messwerten, z.B. ein Rankingverfahren, zulässig oder auch nicht.⁹ Diese Bedingung kann leicht verletzt werden, wenn die zu messende Größe noch von nicht konstant gehaltenen anderen Faktoren abhängt. Die Rendite z.B. kann zwar nicht als Erfolgsmaß für eine einzelne Person, wohl aber für ein Unternehmen gelten, allerdings nur bei vergleichbaren Firmen und für einen hinreichend langen Zeitraum. Die folgenden Aussagen sind Beispiele, in denen die Abhängigkeit von weiteren Größen übersehen wurde. Da unser Blick inzwischen geschärft ist, begnügen wir uns mit der Aufzählung; die Beispiele verdienen jedoch eine genauere Analyse.

*Große Banken können sich darauf verlassen, dass sie gerettet werden, wenn sie in Schwierigkeiten geraten, weil ihr Scheitern zu teuer käme.*¹⁰

*Die politische Passivität steigt, die Partei der Nichtwähler wächst.*¹¹

*Soziologen vermuten einen Zusammenhang zwischen dem sinkenden Interesse an der Politik und der Medienentwicklung. Mehr Information macht offenbar nicht mündiger.*¹²

⁹ In der Psychologie wird diese Problematik unter dem Namen 'Skalentheorie' behandelt. Dass wirklich einmal die zugrundeliegende Skala (= Maßfunktion) bestimmt würde, ist jedoch selten. Ranking und Evaluierung gelten gegenwärtig vor allem im bildungspolitischen Bereich als Heilmittel gegen den Leistungsverfall. Es sind jedoch keine Untersuchungen bekannt, die belegten, dass die zugrundeliegenden Maße die erforderlichen Bedingungen erfüllen.

¹⁰ Allgemeines, inzwischen tief in das Bewusstsein der Bevölkerung eingedrungenes Mantra von Politikern und Bankenvertretern.

¹¹ DEGLER (1993): *Wissenszwerge unter Druck*, p. 156f.

¹² DEGLER (1993): *Wissenszwerge unter Druck*, p. 157.

*Kommunikation besitzt eine die Aggression abbauende und konfliktlösende Kraft. Nichts stabilisiert den Frieden, die Demokratie, die Menschenrechte so sehr wie die freie Kommunikation.*¹³

*Wahlenthaltung ist zumindest in den Demokratien heute ein Ausdruck der Stabilität.*¹⁴

*Sinkende Wahlbeteiligung ist Zeichen einer demokratischen Normalisierung.*¹⁵

Alle Fälle fordern zu der Frage heraus: Woher weiß man das eigentlich? In den beiden letzten Beispielen geht der quantitative Charakter des Gesagten völlig unter: Wahlenthaltung, Stabilität, Normalisierung werden als Ereignis begriffen, das eintreten kann oder nicht; nicht aber, dass es sich um Größen handelt, die von mehr oder weniger großer Intensität sein können. Wären die obigen Äußerungen korrekt, könnte man schließen: Je höher die Wahlenthaltung, desto stabiler bzw. normaler die Demokratie.

Im Alltag begegnet man häufig Fehleinschätzungen, die auf einem unzureichenden Güte- bzw. Gefährlichkeitsmaß beruhen. Vorgesetzte neigen dazu, ihre Leistung zu überschätzen und sich für unentbehrlich zu halten, da sie keine Kritik erfahren und ihnen somit eine Vergleichsmöglichkeit fehlt; unerfahrene Wissenschaftler halten ihre Methode für gut, weil sie keinen Vergleich mit anderen Methoden angestellt haben, und so werden nicht selten Experimente im Labormaßstab auf reale Anwendungsfälle verallgemeinert.¹⁶ Die gleiche Problematik liegt zugrunde, wenn es um die Gefährlichkeit einer Substanz geht, oder, noch allgemeiner, wenn es gilt, die Technikfolgen abzuschätzen, eine Aufgabe, die zunehmend an Bedeutung gewinnt. Die fraglichen Äußerungen haben meist die logisch problematische, unvollständige und zu Fehlinterpretationen herausfordernden Form:

¹³ Äußerungen des Burda-Verlages; zitiert nach DEGLER (1993): *Wissenszwerge unter Druck*, p. 157.

¹⁴ S.M. LIPSET 1962; zitiert nach STARKE (1993): *Krise ohne Ende?*, p. 47.

¹⁵ Diese Ansicht wird kritisiert von STARKE (1993): *Krise ohne Ende?*, p. 48; er gibt auch weitere Einflussfaktoren an.

¹⁶ Diese Haltung ist besonders stark in der Künstlichen Intelligenz ausgeprägt, in der immer gerade das neu aufgekommene Paradigma (etwa Symbolverarbeitung, Produktionssysteme, regelbasierte Verfahren, neuronale Netze) als das einzig Wahre gilt bei gleichzeitiger Verdammung der bereits bekannten Verfahren.

Programm A ist gut.

Substanz A ist krebserregend.

Die Technologie A gefährdet die Menschheit.

Ein Kriterium, wonach sich die behauptete Eigenschaft erkennen ließe, wird meist nicht mit angegeben. Woran erkennt man z.B. die Güte eines Programmes? Um solch ein Kriterium angeben zu können, ist eine Maßfunktion g erforderlich, die den Zusammenhang zwischen einem Leistungsparameter und der Programmgüte beschreibt und somit ein Gütemaß für eine bestimmte Klasse von Programmen darstellt. Sie ermöglicht einen Vergleich zwischen zwei Programmen, denn man kann nun festlegen:

Programm A ist besser als Programm B genau dann wenn $g(p_A) > g(p_B)$;

p_A bzw. p_B sind die gemessenen, ebenfalls nur über ein Maß zugänglichen Leistungen für Programm A bzw. B. Um ein Werturteil abgeben zu können, muß eine Schranke oder ein bestimmtes Gebiet im Zustandsraum ausgezeichnet, d.h. eine Norm angegeben werden. Liefert ein Programm Messwerte, die in dieses Gebiet fallen, wird ihm eine bestimmte Eigenschaft zu-, andernfalls abgesprochen. So könnte man einen bestimmten Leistungsparameter p_0 definieren und über das Gütemaß g festlegen:

Programm A ist gut genau dann wenn $g(p_A) \geq g(p_0)$.

Bei solchen Festlegungen erscheint jetzt das Maßproblem an drei Stellen: bei der Bestimmung von Leistungsparameter und Gütemaß sowie bei der Auszeichnung eines Gebietes, denn selbstverständlich ist letzteres nicht bloß eine Sache der Konvention: Wenn ein Programm das Qualitätskriterium erfüllt, dann soll es auch gut sein: Wie bei Maß und zu messender Größe, so muß auch zwischen ausgewähltem Gebiet und der zugehörigen Eigenschaft eine gesetzmäßige Beziehung bestehen und ebenso wie dort ist auch hier der Nachweis solch einer Beziehung ein Problem, das wenig bewusst und schwierig zu lösen ist.¹⁷ Es bleibt daher meist ungelöst, wie das folgende Beispiel zeigt:

¹⁷ Diese Überlegungen liefern auch eine Antwort darauf, warum ethische Normen bisher dem menschlichen Leben so wenig gerecht geworden sind und warum ihnen häufig etwas abstrakt-weltfremdes anhaftet: Es liegt an den ungültigen Maßbeziehungen, auf denen die Normen beruhen.

Die (wissenschaftlichen) Diskussionen über die "Krise des Parteienstaates", über Politikverdrossenheit und Bürgerunmut waren in den letzten Jahren vor allem dadurch gekennzeichnet, dass zu keinem Zeitpunkt eine Verständigung darüber gelang, wann und aufgrund welcher Beurteilungskriterien von einer Krise gesprochen werden könne. In den meisten Beiträgen ging es stets um einzelne Entwicklungen und Mißstände.¹⁸

Die methodischen Defizite können für die Betroffenen erhebliche Konsequenzen haben. So gehen Unfallopfer nicht selten leer aus, weil sich medizinische Gutachter über die Art und Schwere der Verletzung nicht einigen können.¹⁹ Weitere Alltagsbeispiele sind Einschaltquote und Rendite:

Einschaltquote

Was wird mit der Einschaltquote gemessen? Genauer: Für welche Größe ist die Einschaltquote das empirische Maß?²⁰ In den Medien gilt sie als Maß für die Güte einer Sendung; das ist jedoch unzulässig, denn zwischen Einschaltquote und Güte besteht, wie zahlreiche Beispiele belegen, kein Gesetzeszusammenhang: nicht jede Sendung, die beliebt ist und infolgedessen eine hohe Einschaltquote beschert, muss zugleich auch gut sein. Lässt sich aber nicht wenigstens behaupten, sie sei ein Maß für die Beliebtheit einer Sendung? Doch auch diese abgeschwächte Form lässt sich nicht aufrechterhalten, denn eine Sendung wird bevorzugt relativ zu den gleichzeitig in den übrigen Kanälen gezeigten Programmen. Wegen der Konkurrenz mit anderen Freizeitbeschäftigungen spielt auch die Sendezeit eine Rolle. Einschaltquoten können daher höchstens ein relatives Beliebtheitsmaß sein. In den „Hitlisten der Woche“ werden sie dagegen wie ein absolute Größen behandelt, wobei in einem nachfolgenden Schritt stillschweigend unterstellt wird, eine Sendung sei besser als eine andere, wenn ihre Einschaltquote höher ist als jene. Diese Interpretation ist schon deswegen fragwürdig, weil sie eine bestimmte, niemals nachgewiesene Eigenschaft der Maßfunktion voraussetzt. Einschaltquoten sind daher nichtssagende Werte, denen lediglich

aufgrund von oberflächlichen Eindrücken eine allerdings viel beachtete Bedeutung beigemessen wird.

Rendite

Nach BERTH²¹ ist die Rendite das wichtigste Datum zur Beurteilung des Erfolges von Managern. Sie gilt ihm als ein objektiv Messbares, nachkontrollierbares Resultat, an dem sich der Erfolg auf eine beweisbare Art aufzeigen lässt, *weil Verhalten und Rendite miteinander korreliert werden können.*²² Mit 'korreliert' deutet BERTH offenbar ohne nähere Begründung oder Nachweis einen Gesetzeszusammenhang an, aufgrund dessen es erlaubt wäre zu sagen: 'Die Rendite ist ein Maß für erfolgreiches Managen!'. Das trifft jedoch — m übrigens nach BERTHs eigenen Worten — nicht zu: auf 'kann korreliert werden' lässt sich keine objektive Messung gründen. Soll die Rendite wirklich als Erfolgsmaß gelten, muß sie umkehrbar eindeutig an richtige Entscheidungen gebunden sein. Nun ist gegenwärtig die Rendite der Firmen schlecht, aber nicht, wie man aus dem obigen schließen müsste, wegen schlechten Managements, sondern, wie es heißt, wegen der Rezession.²³ Daher darf man vielleicht vermuten, dass auch eine hohe Rendite weniger dem Verdienst der Manager als vielmehr der guten allgemeinen wirtschaftlichen Lage zuzuschreiben ist. Außerdem ist sie branchenabhängig, so dass Firmen weder hinsichtlich der Rendite gleich, noch aufgrund der Rendite miteinander vergleichbar sind. Wenn aber die Rendite an außerbetriebliche Faktoren gebunden ist, kommt sie als objektives Maß für erfolgreiches Handeln von Einzelpersonen nicht infrage. Offenbar scheint man irrtümlich zu glauben, Berechnungen garantierten automatisch einen Objektivitätsgewinn.

¹⁸ STARKE (1993): *Krise ohne Ende?*, p. 13f.

¹⁹ HAHN (1994) p. 40 — 43. Das im Bewusstsein der Gutachter verankerte Gefühl der Unfehlbarkeit zeigt, dass sich zur Ignoranz nicht selten die Arroganz gesellt.

²⁰ Die Einschaltquote ist selbst ebenfalls nur über ein Maß zugänglich; wir übergehen hier die nicht unproblematische Bestimmungspraxis.

²¹ BERTH (1993): *Erfolg*, p. 23f, p. 42, p. 45.

²² BERTH (1993): *Erfolg*, p. 24.

²³ Wobei freilich eine Rezession als ein Naturereignis angesehen wird, das kommt und geht wie ein Gewitter.

*Gesundheitsregeln*²⁴

Bei den folgenden Beispielen ist stets 'schadet der Gesundheit' hinzuzudenken. Ihr quantitativer Aspekt verbirgt sich in der unklaren Vorstellung, man könne mehr oder weniger gesund sein, deshalb "man sollte etwas für die Gesundheit tun" – zu ergänzen wäre: 'um sie zu verbessern'!

Man muss mindestens drei Liter Wasser am Tag trinken.

Testosteron macht (Männer) aggressiv.

Cholesterin ist schädlich.

Der Körper muss regelmäßig entschlackt werden.

Lesen beim schwachen Licht schadet den Augen.

(a) *Ein Kind sollte möglichst lange gestillt werden.*

Vor dem Sport sollte man sich dehnen.

(b) *Nächtliche Mahlzeiten machen besonders dick.*

Kinderkrankheiten wie Masern bewirken ein besser geschultes Immunsystem.

Krebsfrüherkrankung rettet viele Leben.

Schokolade verursacht Akne.

Rohes Gemüse ist besonders gesund.

(c) *Der Schlaf vor Mitternacht ist der wertvollste.*

(d) *Angeschaltete Handys im Krankenhaus stören die elektrischen Geräte.*

Wir nutzen nur 10 % unseres Gehirns.

Bei chronischen Kopfschmerzen helfen Schmerzmittel.

Quecksilberhaltige Amalgam-Füllungen sollte man entfernen.

Bei Rückenschmerzen hilft Bettruhe.

Weisheitszähne muss man ziehen.

Wir beschränken uns bei unserer Kurzanalyse auf die Fälle (a) – (d):

Fall (a): Ein Kind sollte möglichst lange gestillt werden.

Intendiertes Maß: Je länger die Stillzeit, desto gesünder das Kind X.

Voraussetzungen: Muttermilch ist die perfekte Nahrung für X.
X ist ein Säugling.

Defizite: Es fehlt ein Maß dafür, was mehr oder weniger gesund heißen soll. Das Defizit wird auszugleichen versucht

- durch Hinweis auf Sachzwänge, die der Gesundheit abträgliche sind: keine geeignete Ersatznahrung verfügbar/schlechte hygienische Bedingungen und

- durch zusätzliche Hypothesen wie Stillen beugt Allergien vor

wird Es muss ein Standard *definiert* werden, der angibt, bis wann ein Kind ein Säugling ist (z.B. seine ersten 6 Monate).

Problematisch sind hier die zusätzlichen Hypothesen, deren Wahrheitsgehalt ja erst ermittelt werden muss, bevor sie Teil einer korrekten Schlussfolgerung sein können.

Fall (b): Nächtliche Mahlzeiten machen besonders dick.

Intendiertes Maß: Je später die Mahlzeit, desto gewichtiger wird sie aufgenommen.

Voraussetzungen: Wie Nahrung verwertet wird, hängt bei X von der Tageszeit ab.
„Messergebnis“: Übergewichtige X essen im Vergleich zu normalgewichtigen Y öfter gegen Abend oder in der Nacht.

Bei X löst das Übergewicht keinen besonderen Appetitsschub aus.

X und Y sind Frauen.

Defizite: • Es fehlt ein Maß für den Zusammenhang zwischen Tageszeit und der Intensität der Nahrungsverwertung.

- Ein möglicher Zusammenhang zwischen Körpergewicht und Appetitsschub, der Tageszeit und Nahrungsverwertung entkoppeln würde, bleibt ungeprüft.

²⁴ Übernommen von SCHWEITZER (2014): *Das hilft alles nichts*, p. 29 – 31.

Fall (c):	Der Schlaf vor Mitternacht ist der wertvollste.
Intendiertes Maß:	Je früher die Einschlafzeit, desto besser die Regeneration.
Voraussetzungen:	Die Regeneration hängt von der Einschlafzeit ab. Experimentelles Ergebnis: Spätes Zubettgehen nach Mitternacht beeinträchtigt die Tiefschlafphase. <ul style="list-style-type: none"> • Zusätzliche Hypothese Auf die genaue Zeit nach Mitternacht kommt es nicht an.
Defizite:	Es fehlt ein Maß für mehr oder weniger gute Regeneration. <ul style="list-style-type: none"> • Zusätzliche Hypothese: Die Regeneration ist gut, wenn eine Tiefschlafphase erreicht werden kann.
Fall (d):	Angeschaltete Handys im Krankenhaus stören die elektrischen Geräte.
Defizit:	Mangelnde physikalische Kenntnisse.

Ob solche Ratschläge und Handlungsanweisungen gültig sind, ob sie wirklich helfen, wurde kaum überprüft, vielleicht deshalb, weil sie in der Öffentlichkeit als wissenschaftlich begründet wahrgenommen werden und daher in stillschweigender Übereinkunft schon als überprüft gelten. Über ihre langfristigen gesellschaftlichen Auswirkungen gibt es kaum Untersuchungen; auch hier bewegt man sich im Bereich von mehr oder weniger plausiblen Vermutungen. Dennoch wirken sie! »Sie sagen uns, was wir tun und was wir lassen sollen. Sie engen uns ein. Manche richten sogar Schaden an. Jeder kennt sie, viele verhalten sich danach, sie haben Macht über uns ... Liest man sie dann noch schwarz auf weiß, im Internet, in Ratgeberbüchern, in Zeitschriften ... gibt es kein Entkommen mehr. Und regt sich doch einmal Zweifel, wird er von der inneren Stimme rasch erstickt: Das habe ich so oft gehört, es muss einfach stimmen!«²⁵ Es handelt sich hierbei um die Auswirkungen von Pseudowissen, das im großen Stil medial durch Wiederholungen verbreitet wird.

²⁵ SCHWEITZER (2014): *Das hilft alles nichts*, p. 29f.

4.2 Definierte formale Maße

Mit zunehmender Inkompetenz der Führungskräfte wächst der Bedarf an Maßen, die inhaltliche Bezüge auf formale Kennzeichen reduzieren und sich so ohne jede Sachkenntnis gebrauchen lassen. So gilt z.B. eine Forschungsabteilung für desto erfolgreicher, je mehr Patente sie hervorgebracht hat, und die Anzahl der Veröffentlichungen sowie die Höhe der eingeworbenen Drittmittel ist bei Berufungen inoffiziell das Maß für die Qualität eines Wissenschaftlers. In neuerer Zeit hat der sogenannte Citation index eine besondere Bedeutung erlangt; er beruht auf der Annahme, dass eine Veröffentlichung desto wertvoller (und daher ihr Verfasser desto höher einzustufen) ist, je häufiger sie in renommierten Zeitschriften zitiert wird.²⁶ So wäre die Anzahl der Veröffentlichungen nur dann ein Maß für die Qualität eines Wissenschaftlers, wenn der beste Wissenschaftler auch immer derjenige ist, der das meiste veröffentlicht hat. Formale Maße bestimmen, indem sie in viele Entscheidungen eingehen, zunehmend unser Leben. Man geht mit ihnen recht sorglos um; man verwendet sie, weil sie andere auch verwenden, ohne sich der Fragwürdigkeit und der von ihnen ausgehenden Fehlsteuerung bewusst zu sein. Zu einem gänzlich fiktiven Weltbild führen statistische Größen wie Durchschnitt und Häufigkeit, wenn sie als (formale) Maße für die Bemessung einer Größe verwendet werden.²⁷

Formale Maße werden oft irrtümlich mit dem Argument gerechtfertigt, sie seien objektiver als inhaltliche Bemessungsgrundlagen. Der Vorteil liegt auf der Hand: Man kann eine Abteilung oder einen Bewerber nach "objektiven" Gesichtspunkten einstufen ohne sich mit Patenten und anderen Veröffentlichungen *inhaltlich* auseinandersetzen zu müssen — eine scheinbare probate Methode, von der ja auch in der Physik reichlich Gebrauch gemacht wird, um so die Intensität einer nur schwer oder sogar auf direktem Wege überhaupt unzugänglichen Größe über eine leicht handhabbare in Erfahrung zu bringen. *Strukturell* weichen also beide Ansätze nicht voneinander ab: beide greifen auf Ersatzgrößen zurück, z.B. in der Physik auf einen Zeigerausschlag, um

²⁶ Dieses Maß erfordert ein weiteres, mit dem man die Renormiertheit einer Zeitschrift messen kann; man verwendet hierfür den Impact factor. Über die Problematik dieser Bewertungen siehe KIESER (2003): *Forschung vom Fließband*, p. 31.

²⁷ Z.B. wenn der durch zusätzliche Arbeitstage erwirtschaftete Ertrag - eine völlig virtuelle Größe - die Kosten für die Pflegeversicherung ausgleichen soll.

die elektrische Spannung zu bestimmen. Aber *inhaltlich* besteht ein grundlegender Unterschied zwischen beiden Ansätzen, der es verbietet, sie miteinander zu vergleichen: In der Physik ist zwar die gesuchte Größe gewissermaßen unsichtbar, aber es handelt sich um eine experimentell nachweisbare Größe; sie existiert wirklich. Eine solche Aussage kann man bei einer definierten Größe nicht machen.

Das hat Folgen für die Maßfunktion. Denn was soll es heißen, zwischen einer definierten, d.h. fiktiven Größe und der schließlich gemessenen bestehe ein (natur)gesetzlicher/kausaler Zusammenhang? Kann man denn ein Gesetz, das eine fiktive Größe enthält, überhaupt bestimmen? Offensichtlich nicht: Formale Maße sind daher vom messtheoretischen Standpunkt methodisch unzulässig, sofern sie nicht von einer anerkannten Theorie gedeckt werden.

Weitere Beispiele für nicht-physikalische Messungen sind z.B. Eignungstests und Umfragen. Das Maß ist hier in den Aufgaben und Fragen verborgen, und so müsste nachgewiesen werden, dass diese etwas mit der zu untersuchenden Eigenschaft zu tun haben. In den meisten Fällen legt jedoch lediglich derjenige, der die Tests entwirft, seine ganz speziellen Vorstellungen von der zu messenden Größe in die Aufgaben und Fragen hinein; es wird also keine Beziehung entdeckt, sondern erfunden, z.B. die zwischen Intelligenz(quotient) und dem jeweiligen Intelligenztest. Solch einem durch das Verfahren festgelegtes, definiertes Maß haftet stets eine Willkür an und ist ebenfalls immer problematisch.

Der Gebrauch von falschen Maßen führt zu Fehleinschätzungen der Wirklichkeit und somit zu Fehlentscheidungen. Vor allem die Neigung des Menschen, sich nicht mit einem einmal erreichten Zustand zufrieden zu geben, sondern ihn immer weiter zu optimieren, wirkt sich hier besonders fatal aus: Optimieren bedeutet: sich an irgendeiner Größe ausrichten und so handeln, dass diese Größe entweder möglichst klein oder möglichst groß wird. Sie ist aber oft nicht direkt zugänglich, kann also nicht selbst, sondern nur über die Größe, die für sie als Maß gilt, optimiert werden. Ist es falsch, kommt es zu permanenten, langfristig in die gleiche Richtung wirkenden Fehlsteuerungen; dabei schleifen sich schwer rückgängig zu machende Kreisläufe ein. Ihre Gefährlichkeit beruht auf ihrem stetigen, aber scheinbar unmerklichen Fortschritt. Der vielbeklagte Werteverlust dürfte wohl hierin seine Ursache haben; auch bei den oben erwähnten Beispielen haben sich solche Kreisläufe ergeben:

Das Ziel, nur Sendungen mit möglichst hoher Einschaltquote zu produzieren, führt auf lange Sicht zu immer seichteren Sendungen. Sie prägen den Geschmack des Publikums, und der Geschmack wiederum beeinflusst die Einschaltquoten usw.; u.a. ist daher die Zunahme von Gewaltdarstellung im Fernsehen eine logische Konsequenz dieses Kreislaufs. Wenn allein die Rendite als Erfolgsmaß zählt, werden schlechte Manager auch in guten Zeiten keine Zukunftsinvestitionen tätigen, denn sie schmälern die Rendite. In schlechten Zeiten werden schlechte Manager kurzfristig die Rendite durch Entlassungen und andere Einsparungen heben. In beiden Fällen erzielen sie bessere Ergebnisse als weitsichtige Führungskräfte; auf lange Sicht sind diese Handlungsweisen, wie gegenwärtig zu beobachten, der Ruin jeder Firma. Ein schlechter Wissenschaftler, der für gut gehalten werden möchte, wird nach dem obigen Kriterium sein Ziel erreichen, wenn er es desavouiert, indem er nicht möglichst gute, sondern möglichst viele Publikationen produziert. Das Mengenkriterium führt daher zu einer negative Auslese: Nicht der kreative Forscher, sondern der angepasste Streber, der beamtete Sachverwalter von Wissenschaft, erhält den Vorzug und bildet, seinem Ideal und seinen Fähigkeiten gemäß, Nachwuchs heran.²⁸ Meinungsumfragen wie das Politbarometer sind ein Maß für publikumswirksames Auftreten der Politiker. Diese neigen in fataler Weise dazu, solche Meinungsumfragen überzubewerten, indem sie sich selbst immer häufiger unter das Dauerplebiszit von Umfragen stellen und sich so in ihrer Politik nur noch vom Ausschlag des Meinungspendels leiten lassen.²⁹ Medienwirksamkeit wird angestrebt, nicht die sachliche Lösung von Problemen.

Es gibt eine untrügliche Nagelprobe dafür, ob ein Maß unzulässig ist: Letzteres ist der Fall, wenn die Werte von mindestens einem der Betroffenen (sei es der Experimentator oder die zu untersuchende Institution) beeinflusst werden können.

²⁸ Beispiele für die Anpassung an die Bewertungskriterien ‚citation index‘ und ‚impact factor‘ siehe KIESER (2003): *Forschung vom Fließband*, p. 31.

²⁹ STARKE (1993): *Krise ohne Ende?*, p. 65.

4.3 Score-Methoden und Ranking

Nun hat sich aber herausgestellt, dass eine theoretische Größe wie 'Innovation' oder 'Wirtschaftswachstum' von mehreren Faktoren abhängt, also nicht durch eine einzige Maßzahl zureichend erfasst werden kann. Man versucht deshalb, dieses grundsätzliche Problem durch einen Score in den Griff zu bekommen.

In einen „Score“ S gehen stets $n > 1$ Messeigenschaften (Indikatoren) Z_1, \dots, Z_n ein, deren Intensitäten z_1, \dots, z_n mit g_1, \dots, g_n gewichtet und dann linear verknüpft werden. Die Maßfunktion ist dann ein gewichtetes arithmetisches Mittel der Form

$$S = g_1 z_1 + g_2 z_2 + \dots + g_n z_n, \quad g_1, \dots, g_n > 0.$$

Damit stellt ein Score ebenfalls ein definiertes formales Maß dar und ist schon aus diesem Grund immer willkürlich. Als „multidimensional“ definierte Größe bereitet es jedoch noch zusätzliche Probleme:

- (1) Die Messeigenschaften Z_1, \dots, Z_n wählen wenige, sich auf ihre Erfahrung berufene Personen aus; die Auswahl ist daher immer subjektiv und oft kultur- und/oder landesspezifisch.
- (2) Die Werte der Gewichte g_1, \dots, g_n werden vordergründig "erraten"; sie können aber bestenfalls statistisch abgeschätzt werden.
- (3) Naturgesetze sind nicht additiv verknüpft; es bleibt daher unklar, welchen Wert überhaupt solch eine Verknüpfung besitzt. Die für die Interpretation so wichtige strenge Monotonie geht durch die Additionen verloren: Im allgemeinen Fall hat der Score S geometrisch die Form einer uneindeutigen „Bergoberfläche“.
- (4) Die Linearitätsannahme ist willkürlich; es könnten ja z.B. auch quadratische Terme mit in den Score eingehen.

- (5) Die einzelnen den Score definierenden Institute legen jeweils ihre eigene Score-Gleichung zugrunde; das führt zu unterschiedlichen, nicht miteinander vergleichbaren Score-Werten.

Fazit: Ein Score, allgemein: jede Maßzahl x , welche durch

$$x = g_1 z_1 + g_2 z_2 + \dots + g_n z_n$$

berechnet wird, hat keinen wissenschaftlichen Wert und ist daher auch kein objektives Maß (wie oft behauptet wird). Insbesondere erlaubt es keine Kleiner/Größeraussagen, denn solche Aussagen sind nur für eindimensionale Fälle möglich, aber diese sollten eben gerade durch Verwendung eines Score überwunden werden. Außerdem kann man aber bei Bewertungen auf Kleiner/Größeraussagen nicht verzichten, denn schließlich möchte man z.B. sagen können, die Innovationsleistung eines Landes sei größer als die in einem anderen oder eine Institution habe seine Innovationsleistung verbessert. Im Verfahren selbst ist also bereits ein unlösbarer Widerspruch verankert; er muss sich in der Vergeblichkeit äußern, das Verfahren zu verbessern. Hierfür gibt es in der Tat erste Anzeichen: Nach SCHEFOLD & Lenz »scheint das Leitbild „Wissensgesellschaft“ vor allem dem Zweck zu dienen, die Bedeutung von Wissenschaft und Technologie für die wirtschaftliche Entwicklung in politischen Programmatiken hervorzuheben und daran anknüpfend den Stellenwert empirischer Indikatoren zur Innovationsleistung und Wissensbasierung wirtschaftlicher Strukturen im Rahmen der europäischen Evaluierungskultur zu erhöhen. Dabei ist zu beobachten, dass diese Entwicklung keine wesentlichen Fortschritte in der empirischen Analyse nach sich gezogen hat, so dass lediglich bereits existierende Statistiken und herkömmliche Input-/Output-Indikatoren unter dem neuen Label „knowledge economy“ zusammengefasst werden ... Auch die zahlreichen Anstrengungen von EUROSTAT oder der OECD zur Fortentwicklung empirischer Indikatoren und Methoden haben bisher nicht die erwünschten Resultate erzielt.«³⁰ Wenn unsere Analyse zutrifft, ist man zur Zeit bei EUROSTAT, OECD und anderen Instituten dabei, gewissermaßen das perpetuum mobile für Score-Verfahren zu erfinden.

³⁰ SCHEFOLD (2008): *Einleitung*, p. 6; (Sperrung PJ).

5 Qualitative versus quantitative Beziehungen

Die bisherigen Ausführungen könnten irrtümlicherweise den Eindruck erwecken, die angesprochenen Probleme betreffen lediglich quantitative Aspekte, die auf der begrifflichen Ebene ohne Belang seien. Doch auch die Prädikate im Sinne der Prädikatenlogik, z.B. die Beziehung zwischen zwei Individuen ('A ist Sohn von B'), die Zuordnung eines Elementes zu einer Klasse ('A gehört zu B') oder die Behauptung von Gemeinsamkeiten ('Alle A haben die Eigenschaft B') usw., die als rein begriffliche Beziehungen gelten, fußen auf Gesetzmäßigkeiten, so dass es nicht ins Belieben gestellt ist, Merkmale zu definieren und danach Objekte zu klassifizieren. Die Merkmale eines Säugetierskeletts z.B. können nur dann als solche gelten, wenn sie tatsächlich bei allen Säugetieren vorkommen und nicht bloß definiert wurden. Die Gesetzmäßigkeit besteht hier darin, dass es gemeinsame Merkmale gibt. Auch wenn man also nicht bis zur Formulierung des zugrundeliegenden Gesetzes vordringt, sondern sich auf die begriffliche Beschreibung von Beziehungen beschränkt, hat man es im Hintergrund mit Prozessen zu tun, denen man diese Beziehungen verdankt. Die Forderung, die Maßbeziehung zu klären, besagt daher nichts anderes als den Kausalzusammenhang nachzuweisen; dies erfordert quantitative Methoden.

Umgekehrt gehört auch zu den quantitativen Beziehungen ein begriffliches Pendant. 'Die Zeitdauer, die der Schall unterwegs ist, ist ein Maß für die Wassertiefe', oder allgemein: 'A ist Maß von B', drückt eine begriffliche Beziehung aus. Der quantitative Aspekt wird hier ebenso wenig sichtbar wie in den obigen Fällen, obwohl durch diese Aussage eine Wechselwirkung postuliert wird, die dafür sorgt, dass sich die Intensität der beiden Größen verändert. Da eine Wechselwirkung voraussetzt, dass die betreffenden Wechselwirkungspartner verschiedene Zustände einnehmen können, enthält eine solche Behauptung stets einen quantitativen Bezug, ein Mehr oder Weniger, in dem das Kräfteverhältnis der wechselseitigen Abhängigkeit erfasst ist; es wird durch die Maßfunktion beschrieben. D.h. die Beziehungen zwischen Größen — sie stellen die Bedingung für die Möglichkeit von Wissen dar — kommen nur dadurch zustande, dass es Systeme gibt, die mehr als einen Zustand einnehmen. Die Welt der Begriffe, von der quantitativen Welt abgetrennt, ist eine tote Welt; sie geht von Aussagen aus und gibt an, welche Folgerungen sich ergeben, wenn diese Aussagen wahr sind. Warum diese Fakten wahr sind, wie sie zustandekommen, darüber erfahren wir nichts. Ebenso wenig ist eine Welt ohne Begriffe denkbar: Quantitative und

qualitativ-begriffliche Aspekte bedingen einander.³¹ Jede einseitige Betonung birgt die Gefahr einer folgenreicheren Desorientierung, so z.B. beim sogenannten offenen Kodieren in der Qualitativen Forschung.³²

6 Resümee

Mißverständnisse und Fehlentscheidungen haben oft ihre Ursache in einer mangelnden Realitätserfassung: Maßregeln werden getroffen, ohne sich ein korrektes Bild von der Wirklichkeit zu machen; selbst offensichtliche Misserfolge werden häufig zu spät oder gar nicht erkannt. In Konferenzen, Vorträgen und Veröffentlichungen nimmt die Propagierung von Lösungsvorschlägen und neuen Methoden einen breiten Raum ein, während das zu lösende Problem, ein Stiefdasein führt und meist unerwähnt bleibt; auch dies ist ein Zeichen von Realitätsverlust.

Einen Einblick in die realen Verhältnisse gewinnen, wurde als ein Vorgang betrachtet, der in irgendeiner Weise mit Messung verbunden ist. Messen, wenn es den Namen verdienen soll, muß nach bestimmten Prinzipien erfolgen. Eines der grundlegendsten von ihnen, das Maßprinzip, haben wir herausgegriffen. Wir haben an mehreren Beispielen gezeigt, dass hinsichtlich dieses offenbar weitgehend unbekanntem Prinzips wenig Problembewusstsein entwickelt ist, denn es wird oft in gravierender Weise verletzt. Ebenso wie es nach WATZLAWICK nicht möglich ist, sich nicht zu verhalten, so ist es auch nicht möglich, Maße und Maßbeziehungen zu umgehen: Selbst wenn man sie nicht angeben kann oder sich ihrer nicht bewusst ist, liegen sie doch jeder Ermittlung des Istzustandes zugrunde und führen, sofern sie unzutreffend sind, zu zweifelhaften, Desorientierung und Fehlentwicklungen auslösenden Ergebnissen.

Die Desorientierung entsteht in diesem Fall nicht aufgrund falscher Überlegungen oder eines fehlerhaften Gedankengebäudes, sondern allein aufgrund unzulässiger Ausgangsdaten; sie führt, eine Ebene höher, zu einer weiteren Desorientierung: Man hält die getroffene Entscheidung für

³¹ Und so ist das eine nicht etwa eine Vorstufe des anderen, wie in der wissenschaftstheoretischen Literatur behauptet wird, siehe "Von der Qualität zur Quantität. Intuitiv-konstruktive Theorie der wissenschaftlichen Begriffsformen" in: STEGMÜLLER (1970): *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie Bd. II*, p. 15 — 105.

³² JAENECKE (2013): Offenes Kodieren in der Qualitativen Forschung.

theoretisch fundiert, doch da die erhoffte Wirkung ausbleibt, die Überlegungen aber korrekt waren, muß, so schließt man, die theoretische Grundlage falsch gewesen sein. Auf diese Weise kommt die Theorie und die theoretische Arbeit in Verruf und es entsteht die Meinung, die wissenschaftlich-analytischen Methoden würden versagen. Tatsächlich liegt hier aber ein methodischer Fehler vor: Man trennt nicht die Theorie und die empirische Bestimmung des Istzustandes, der zur Anwendung der Theorie gebraucht wird, sondern will ihn ebenfalls durch theoretische Überlegungen gewinnen: Man theoretisiert dort, wo eine empirische Vorgehensweise gefordert ist; man verhält sich so wie ein Richter, der sich nicht bei den Beteiligten ein Bild über den Tatbestand macht, sondern versucht, ihn aus dem Strafgesetzbuch abzuleiten.

Auch das Fehlurteil in die andere Richtung ist weit verbreitet: Weil die Daten mit wissenschaftlichem Rüstzeug bestimmt wurden, hält man sie für objektiv; der technische Aufwand bei einer Messung gilt als Maß für Objektivität, eine Ansicht, die sich oft in Formulierungen wie: 'wissenschaftliche Untersuchungen haben ergeben' kleidet; man bedenkt nicht, dass wissenschaftliche Methoden auch falsch angewendet werden können. So liegt bei der Beurteilung der Zuverlässigkeit von Datenmaterial selbst ein falsches Maß zugrunde.

In den Wissenschaften ist das methodische Bewusstsein im Niedergang begriffen; dies zeigt sich u.a. in dem sorglosen Umgang mit Maßen und der geringen Reflexionstiefe der Argumente. Gefäuscht durch gefühlmäßige Plausibilität und den offensichtlichen Anschein sieht sich nur selten jemand veranlasst, ein Maß in Frage zu stellen oder Aussagen auf tieferliegende quantitative Beziehungen hin zu durchleuchten. Man verkennt dabei das dieser Arbeit innewohnende schöpferische Potential, gilt es doch, Gesetzmäßigkeiten aufzuspüren und den Grundstein für eine solide Begriffsklärung zu legen. Solche Gesetzmäßigkeiten könnten überdies ein erstes Ansatz für die Formulierung einer Theorie sein, auf einem Gebiet, in dem derartige Versuche bisher gescheitert sind.

7 Anhang

Desorientierungen durch falsche Maße finden sich in reichem Maße in Themen von gesellschaftspolitischer Bedeutung. Sie sind oft unauffällig und werden verdeckt durch unreflektierte Erwartungshaltungen und ideologischen Vorstellungen, gleichwohl können sie tief in das Leben der Menschen eingreifen und Schicksal spielen; dies belegt in eindringlicher und zugleich in beunruhigender Weise die folgende Beispielsammlung.

7.1 Score-abhängige Beispiele

»Der Begriff "Kultur" bezieht sich auf Werte, Normen, Verhaltensweisen oder Traditionen, also immaterielle Phänomene, die das Verhalten und die Lebensweise von Menschen bestimmen. Wenn überhaupt, so können solche Werte nur über definierte Maße erschlossen werden.«³³ So ersucht RICHARD FLORIDA den Zusammenhang zwischen kulturellen Orientierungen und technologischer Leistungsfähigkeit aufzuzeigen, wobei er sich auf einen indikatorenbasierten Messansatz stützt.]

7.1.1 Messkonzepte für kulturelle Leistungen

Score-abhängige Verfahren dienen heute dazu, die Leistungsfähigkeit von Personen oder Institutionen zu bemessen. Am einfachsten wäre es, wenn sich Personen und Institutionen auf eine einzige Eigenschaft reduzieren ließen. Dann könnte man deren Intensität bestimmen und sie der Bewertung zugrundelegen. Doch dabei erfasste man jedoch nur eine einzige Leistungskomponente, die nicht in der Lage ist, das gesamte Leistungsspektrum widerzuspiegeln; es müssen also stets $n > 1$ Indikatoren x_1, x_2, \dots, x_n berücksichtigt werden, aus deren n Intensitäten x_1, x_2, \dots, x_n in geeigneter Weise ein Gesamtwert, der Score, zu bilden ist. Hinter der Redewendung 'in geeigneter Weise' verbirgt sich die Vorstellung, man müsse diese Werte, um Objektivität zu erreichen, nur in einer geeigneten Weise gewichten; es seien dies die Gewichte g_1, g_2, \dots, g_n . Der Score S ergibt sich dann aus der additiven Verknüpfung

³³ SCHMOCH (2008): *Messkonzepte für Innovationskultur*, p. 99.

$$(1) \quad S = g_1 x_1 + g_2 x_2 + \dots + g_n x_n.$$

Angenommen, es gäbe Messverfahren zur Bestimmung der Intensitäten x_1, x_2, \dots, x_n , so bleibt doch noch die Frage, auf welche Weise sich die Gewichte g_1, g_2, \dots, g_n zu gewinnen sind? Streng genommen müssten sie auch durch Messungen bestimmt werden, aber das setzt voraus, dass man die Beziehungen zwischen den einzelnen Indikatoren feststellen könnte; hierfür gibt es aber kein anerkanntes Verfahren: Die Score-Gleichung (1) kann, wie immer man sie auch interpretiert, aufgrund der folgenden Mängelliste keine Objektivität garantieren:

- Die Indikatoren x_1, x_2, \dots, x_n wählen wenige, sich auf ihre Erfahrung berufene Personen aus; die Auswahl ist daher immer subjektiv und oft kultur- und/oder landesspezifisch.
- Die Werte der Gewichte g_1, \dots, g_n werden vordergründig "erraten"; sie können aber bestenfalls statistisch abgeschätzt werden.
- Naturgesetze sind nicht additiv verknüpft; es bleibt daher unklar, welchen Wert solch eine Verknüpfung besitzt: wie lässt sich bei einer additiven Verknüpfung ein streng monotonen Verhalten realisieren?
- Die Linearitätsannahme ist willkürlich; es könnten ja z.B. auch quadratische Terme mit in den Score eingehen.
- Die einzelnen Institute, welche das Ranking vornehmen, legen jeweils ihre eigene Score-Gleichung zugrunde; das führt zu unterschiedlichen, nicht miteinander vergleichbaren Score-Werten.
- Eine Fehlerabschätzung ist offenbar für Score-Werte nicht vorgesehen; doch ist ein Messwert ohne Fehlerabschätzung eine wertlose Zahl, aus der man nichts schließen kann. Er ist ohne solch eine Abschätzung sinnlos.

Man wird jedoch z.B. kaum Wirtschaftsdaten finden, bei denen eine Fehlerabschätzung angegeben ist. Dennoch werden aus solchen Daten Schlußfolgerungen gezogen. So spricht man z.B. von einem wirtschaftlichen Aufschwung, obwohl die zugrundegelegten Daten so wenig von Null

abweichen, dass man berechtigterweise annehmen darf, dass sie noch innerhalb der Messgenauigkeit liegen.

Es ist sehr zweifelhaft, ob sich diese Willkürlichkeiten grundsätzlich beheben lassen. Solange sie jedoch bestehen, müssen Score-abhängige Verfahren als wissenschaftlich wertlos eingestuft werden. Verwendet man dennoch Score-Werten als Entscheidungsgrundlage, so sind Fehlentscheidungen die unvermeidliche Folge.

7.1.2 Fachspezifisches Ranking und fachübergreifendes Hochschul-Ranking

Bekannt seien k zu einer Bewertung vorgesehene Personen oder Institutionen. Bei einem Ranking-Verfahren bestimmt man zuerst deren Score-Werte S_1, S_2, \dots, S_k gemäß Gleichung (1) und ordnet sie anschließend der Größe nach; man erstellt also eine Rangreihenfolge in aufsteigender Richtung, wobei die oberen Plätze in der Reihe zugleich die Sonnenplätze sind. Beide Ranking-Varianten folgen dem in Gleichung (1) beschriebenen Muster; sie unterscheiden sich lediglich in der Auswahl der als maßgeblich zugrundegelegten Indikatoren.

7.1.3 Global MBA-Ranking

Das wohl wichtigste Bewertungsverfahren in der Wirtschaftswissenschaft ist das Global MBA-Ranking; 'MBA' hier steht für 'Master of Business Administration'. Die zur Berechnung herangezogenen Indikatoren, auf die ja alles ankommt, werden meist umgangssprachlich und unvollständig beschrieben. Um es z.B. im MBA Ranking der *Financial Times (FT)* nach oben zu schaffen, muss es einer Business School vor allem gelingen, ihre Absolventen in gut bezahlte Jobs zu vermitteln. Das durchschnittliche Gehalt und die Gehaltssteigerung im Vergleich zu der Zeit vor dem MBA Studium machen jeweils 20 Prozent der Bewertung aus. Neben dem Verdienst ist die Forschungsleistung das wichtigste Kriterium: Die *FT* zählt, wie viele Fachartikel Professoren und wissenschaftliche Mitarbeiter in 45 ausgewählten Fachzeitschriften veröffentlicht haben.³⁴ Extrahiert man hier die einzelnen Komponenten, so erhält man als Kriterien

³⁴ BUHSE (2014): *Die Macht der rosa Tabelle*, p. 85.

- Durchschnittliches Gehalt (X_1).
- Gehaltssteigerung im Vergleich zu der Zeit vor dem MBA Studium (X_2)
- Anzahl der Fachartikel in 45 ausgewählten Fachzeitschriften (X_3).

Bleibt es, wie im Folgenden angenommen, bei diesen drei Komponenten und der erwähnten 20% - Bewertung, dann ergäbe sich die Score-Gleichung

$$S = 0,2x_1 + 0,2x_2 + 0,6x_3.$$

Der Vorteil dieser Verfahrensweise liegt auf der Hand: Die Bewerter brauchen sich nicht mit Inhalten zu befassen; entschieden wird rein „formal“, und so entscheiden ein paar schöne Tabellen über Ruhm und Ehre in der Welt der Business Schools.³⁵ Dabei werden allerdings stillschweigend drei Maßfunktionen installiert:

Die Güte einer Schule ist desto höher,

- je größer das Gehalt ihrer Absolventen ist (x_1);
- je größer die Gehaltssteigerung nach dem Studium ausfällt (x_2) und
- je mehr Veröffentlichungen die Schulangehörigen in ausgewählten Zeitschriften publiziert haben (x_3).

Es sei daran erinnert, dass eine Maßfunktion nur ihren Auftrag, nämlich Maß für eine andere Größe zu sein, erfüllen kann, wenn sie einen kausalen, also gesetzmäßigen Zusammenhang erfasst, der ohne Ausnahme immer gilt. Keine der obigen Maßfunktionen wird diesem Anspruch gerecht. Um das nachzuweisen, genügt die oben erwähnte Nagelprobe: Ein Maß ist immer unzulässig, wenn die Werte, von dem, der beurteilt werden soll, beeinflusst werden können. Im obigen Beispiel sind Gehalt und Gehaltssteigerung verhandelbar, und die Anzahl der Veröffentlichungen lassen sich leicht durch subtile Methoden steigern. Überdies erhöht charakterloses Handeln die vergütungssteigernden Aufstiegschancen.

Was sind die Folgen? Hier gilt es zu unterscheiden zwischen den unmittelbaren Folgen und solchen, die sich aus einer langjährigen Entwicklung ergeben. Wer etwa bei den jährlich veröf-

³⁵ BUHSE (2014): *Die Macht der rosa Tabelle*, p. 85.

fentlichten Ranking-Werten der Wirtschaftszeitung *Financial Times* »gut abschneidet, bekommt die besten Studenten, die lukrativsten Forschungsaufträge und kann hohe Studiengebühren verlangen.«³⁶ Das sind die unmittelbaren Folgen. Wendet man dieses Verfahren jedoch über Jahre hinaus an, so gelangt man in einen sterilen Kreislauf: Wenn ein Absolvent von einer Schule mit einem hohen Score-Wert kommt, dann fühlt sich eine Personalabteilung veranlasst, diesen Tatbestand in einer höheren Vergütung zum Ausdruck zu bringen. Diese höhere Vergütung geht dann aber in die Bewertung des nächsten Jahres mit ein und hilft damit der Schule, sich wiederum ihren oberen Rangplatz zu sichern. Es ist daher nicht verwunderlich, dass in den 2014 veröffentlichten Ranking-Werten der *Financial Times* die großen Hochschulen aus den USA dominieren. Unter den Top 10 sind sieben US-Unis.³⁷

Langfristig produziert solch ein Kreislauf eine negative Auslese: Nicht diejenigen, die wirklich in der Lage wären, eine Krise zu meistern, gelangen in Führungspositionen, sondern diejenigen, deren Schule die höchsten Reputationen besitzt. Dass sich daraus fatale gesellschaftliche Konsequenzen ergeben können, haben uns die jüngsten von „Experten“ eingebrockten Finanzkrisen vorgeführt. Im Übrigen ist in den Veröffentlichungen auch der renommierten Zeitschriften ein unablässiger Niedergang des methodischen Bewusstseins zu beobachten. Wenn aber die Score-Werte tatsächlich etwas über den jeweiligen wissenschaftlichen Stand einer Schule aussagten, müssten sich eigentlich allein durch das Bestreben, den Score-Wert zu erhöhen, automatisch immer gehaltvollere Veröffentlichungen ergeben.

7.1.4 CHE-Ranking

Was für den speziellen Fall des BSA-Rankings gilt, trifft auch sinngemäß auf andere Verfahren zu, die andere Fachverbände oder Hochschulen als Ganzes zu bewerten versuchen, wie dies beim Hochschul-Ranking des Centrums für Hochschulentwicklung (CHE) geschieht. Gegen solche Ranking-Ansätze haben sich in Deutschland Fachverbände gewehrt: »Die Tatsache, dass bereits etliche Fachverbände wie etwa der Soziologie oder der Geschichte ihre Mitglieder aufgefordert haben, aus dem Ranking – endgültig! – auszusteigen, belegt in eindeutiger Weise, dass die

³⁶ BUHSE (2014): *Die Macht der rosa Tabelle*, p. 85.

³⁷ BUHSE (2014): *Die Macht der rosa Tabelle*, p. 85.

Kritik am Messkonzept des CHE breit verankert ist.«³⁸ Diese Kritik ist aber durchaus berechtigt, und wird auch teilweise eingestanden: Soweit bekannt, gibt es bislang noch immer keine vollständige Liste der Indikatoren samt ihrer Gewichte, mit denen man eine Score-Gleichung aufstellen könnte, jedenfalls keine öffentlich zugängliche. Es waltet eine gewisse Geheimniskrämerei, auch gegenüber möglichen Erweiterungen der Berechnungsmethoden: »Kritik an den Messmethoden wurde regelmäßig nicht zur Kenntnis genommen.«³⁹

Nach Ansicht von ABELS & KNEUER falle es in den Aufgabenbereich der Politikwissenschaft, »sich mit diesen Defiziten zu beschäftigen; alles andere wäre ein wissenschaftliches Armutzeugnis. Was für ein Qualitätsausweis wäre das, wenn eine Wissenschaft, zu deren Lehr- und Forschungsinhalten Messmethoden gehören, fehlerhafte Rankings widerspruchlos hinnehmen würde? Es geht allein um das korrekte Messen.«⁴⁰

Durch Gründung eines Moderatoriums glaubte man die Gesprächsbereitschaft des CHE erhöhen zu können und hofft nun, bei einem erfolgreichen Gesprächsverlauf im Jahr 2015 ein solides Ranking zur Verfügung zu haben.⁴¹ Wie schon aus der Formulierung "es geht allein um das korrekte Messen" hervorgeht, hält man die derzeitigen Methoden für ungenügend, und alle Hoffnungen ruhen auf der trügerische Annahme, man müsse lediglich nach besseren Methoden suchen, dann ließe sich auch der Mangel beheben. Nun spricht man in den Sozialwissenschaften häufig von Messmethoden, aber die wissenschaftstheoretisch orientierte Messtheorie, deren Aufgabe es ist, sich mit den Messmethoden zu befassen, wurde bislang kaum zur Kenntnis genommen. Man hätte sonst erfahren, dass die Mängel nicht aus unzulänglichen Methoden erwachsen, sondern bereits im widersprüchlichen Ansatz stecken, auf denen die Methoden beruhen. Die oben angegebene Mängelliste kann offenbar grundsätzlich nicht behoben werden, hat man doch noch nicht einmal ein anerkanntes objektives Maß, wonach man entscheiden könnte, welche von zwei Methoden die bessere ist. Wie will man ohne solch ein Maß, also ohne ein rechtes Ziel zu haben, zum Erfolg kommen? Es ist abzusehen, dass sich von den zahlreichen Expertenmeinungen am Ende einige wenige durchsetzen werden und diese werden in einen neuen Satz von

Indikatoren einfließen, aber ob sich dadurch die Lage wirklich verbessert, muss in Ermangelung eines geeigneten Maßes offenbleiben. Eines ist aber gewiss: die Score-Werte werden vielleicht nicht von besserer Qualität, aber auf jeden Fall anders sein.

7.1.5 Shanghai-Ranking

Massive Kritik am Hochschul-Ranking ist nicht neu. Im kontinentalen Europa ausgelöst wurde sie vermutlich durch das mit viel Geheimnistuerei verbundene Shanghai-Ranking von 2003, bei dem vor allem deren Universitäten schlecht abschnitten, während die us-amerikanischen und britischen Universitäten die Spitzenplätze besetzen. Bei der medialen Berichterstattung wurde nicht auf Seriosität geachtet; im Vordergrund standen wie so oft die „alarmierenden“ negativen Nachrichten. Dabei wäre es gar nicht so schwierig gewesen, das Anstößige zu benennen, denn man trifft immer auf die gleichen gravierenden messtheoretischen Mängel. »Es sind nur eine Handvoll Daten, nach dem Gusto der Forscher zusammengestellt, die über Gewinner und Verlierer entscheiden. Ganze sechs Forschungskennzahlen definieren laut Shanghai-Ranking, wer sich zu dem erlesenen Klub der international führenden Universitäten rechnen darf ... Die Lehre fällt dabei durchs Raster, auch sonst gilt: Wenn eine Uni nicht den Vorstellungen der Ranking-Macher von Exzellenz entspricht, hat sie verloren.«⁴² Zu den sechs entscheidenden Indikatoren »zählen die Anzahl der Nobelpreisträger, die Anzahl der von Uni-Forschern in den englischsprachigen Magazinen *Nature* und *Science* veröffentlichten Artikel und die Häufigkeit der wissenschaftlichen Zitationen durch Dritte: ein großer Vorteil für etablierte Forschungsuniversitäten.«⁴³ Auch hier zeigt sich erneut die sich selbstverstärkende, eine Objektivität vortäuschende positive Rückkopplungsschleife zugunsten einiger weniger Etablierter.

Die geäußerte Kritik an den unzulässigen Methoden leidet jedoch selbst unter messtheoretischem Methodenmangel; die eigentlichen Mängelursachen werden nur lückenhaft erkannt und in die Argumente fließen häufig altbekannte vordergründige Ressentiments. Ein faires internationales Ranking ist unmöglich behaupten jene Wissenschaftler, die schon nationale Hochschulvergleiche ablehnen. Sie haben aus messtheoretischen Gründen Recht. »Sie befürchten, dass willkür-

³⁸ ABELS & KEUNER (2013): »Es geht allein ums korrekte Messen«, p.69.

³⁹ ABELS & KEUNER (2013): »Es geht allein ums korrekte Messen«, p.69.

⁴⁰ ABELS & KEUNER (2013): »Es geht allein ums korrekte Messen«, p.69.

⁴¹ ABELS & KEUNER (2013): »Es geht allein ums korrekte Messen«, p.69.

⁴² WIARDA (2009): *Auf die Plätze, fertig – los*, p. 69.

⁴³ WIARDA (2009): *Auf die Plätze, fertig – los*, p. 69.

lich gesetzte Effizienzkriterien die Forscher in ihrer Freiheit beschränken könnten. Auch kritisieren sie [ebenfalls zu Recht] eine Verzerrung des Wettbewerbs: Diejenigen Hochschulen, die in den Vergleichen gut abschneiden, bekämen als Belohnung mehr Geld vom Staat – und könnten die andere dann erst recht abhängen.«⁴⁴

Als einzigen Ausweg aus der Krise sieht man wiederum nur die Verbesserung in den Methoden. So hat das CHE die multidimensionale Methode eingeführt, welche bis zu 34 Kriterien vorsieht, und zwar auf der Basis einzelner Fächer, niemals ganzer Hochschulen.⁴⁵ Es fand also eine Verfeinerung statt, wobei offenbar unbeachtet blieb, dass mit der steigenden Anzahl der Kriterien auch die Probleme anwachsen könnten, z.B. die, welche beim dem Versuch entstehen, die Gewichte zu ermitteln. Im Auftrag der EU-Kommission soll ein multidimensionales und globales Hochschulranking entwickelt werden. Dabei werden zunächst (») die Hochschulen entsprechend ihrer Struktur in mehrere Typen unterteilen, damit nicht mehr internationale Forschungsinstitute mit regionalen Fachhochschulen, Volluniversitäten mit Spartenanbietern verglichen werden. Dann wird die im nationalen Ranking eingeführte Methodik so weiterentwickelt, dass es für die verschiedenen Hochschultypen auch unterschiedliche Typen gibt. Bei den forschungsintensiven Universitäten werden die Publikationen stärker zählen als bei regionalen.«⁴⁶

Hochschulvergleiche durch eine europaweite Uni-Datenbank. Abrufbar sind Budget, Zahl der Studenten und Professoren, mögliche Abschlüsse, Forschungsfelder usw.

Nur selten wird in der Literatur gefragt, welche Ziele mit einem Hochschul-Ranking erreicht werden sollen. Nach ZIEGELE, Leiter des CHE, richtet sich das CHE-Ranking »an mehrere Gruppen. Zum einen geben wir Abiturienten ein Instrument in die Hand, um die Studienangebote der Universitäten anhand verschiedener Kriterien zu vergleichen. Zudem helfen wir der Politik und den Hochschulleitungen, rationale Entscheidungen zu treffen, die auf Daten basieren und nicht auf subjektiven Einschätzungen.«⁴⁷

⁴⁴ WIARDA (2009): *Auf die Plätze, fertig – los*, p. 69.

⁴⁵ WIARDA (2009): *Auf die Plätze, fertig – los*, p. 69.

⁴⁶ WIARDA (2009): *Auf die Plätze, fertig – los*, p. 69.

⁴⁷ SPIEWAK (2010): »Sie führen die Studenten in die Irre«, p. 69.«

8 Literatur

- ABELS, GABRIELE & KNEUER, MARIANNE (2013): »Es geht allein um korrektes Messen.« Warum die Politikwissenschaft beim Uni-Ranking aussetzt – eine Replik. *DIE ZEIT* 26, 20.06.13, p. 69.
- BERTH, R. (1993): *Erfolg. Überlegenheitsmanagement: 12 Mind-Profit Strategien mit ausführlichem Testprogramm*. Econ Verlag, Düsseldorf etc. 1993.
- BUHSE, MALTE (2014): Die Macht der rosa Tabelle. Wer es als Business School in das Ranking der »Financial Times« schafft, ist oben angekommen. *DIE ZEIT* 40, 25.09.14, p. 85.
- DEGLER, H.-D. (1993): *Wissenszweige unter Druck. Der Spiegel* 1993, Heft 4.
- GRUPP, HARIOLD (2008): Kommentar zum Beitrag von Ulrich Schmoch: Quantifizierung des Qualitativen? In: SCHEFOLD, BERTRAM & LENZ, THORSTEN [Hrsg.] (2008): *Europäische Wissenschaftsgesellschaft*, p. 113 – 115.
- GRUPP, HARIOLD (2008): Innovationsökonomische Perspektiven der Wissenschaftsgesellschaft. In: SCHEFOLD, BERTRAM & LENZ, THORSTEN [Hrsg.] (2008): *Europäische Wissenschaftsgesellschaft*, p. 19 – 28.
- HAHN (1994): *ADAC Motorwelt* 1994, Heft 2, p. 40 – 43.
- JAENECKE, PETER (1982): Grundzüge einer Meßtheorie. *Zeitschrift für Allgemeine Wissenschaftstheorie* 13, 1982, p. 234 – 279.
- JAENECKE, PETER (2013): Offenes Kodieren in der Qualitativen Forschung. Im Internet verfügbar unter: <http://www.peterjaenecke.de/soziologie.html> .
- JAENECKE, PETER (2014): Einführung in die Messtheorie. Im Internet verfügbar unter: <http://www.peterjaenecke.de/messtheorie.html> .
- KIESER, ALFRED (2003): Forschung vom Fließband. Evaluation schafft erst die Wirklichkeit, die zu bewerten sie vorgibt. *DIE ZEIT* 30, 17.07.2003, p. 31. Im Internet am 09.10.14 verfügbar unter: <http://www.zeit.de/2003/30/B-Zitate>
- LENZ, THORSTEN & SCHEFOLD, BERTRAM (2008): *Einleitung: Die „Europäische Wissenschaftsgesellschaft“*, p. 1 – 16.
- NEUBEISER, M.-L. (1993): *Die Logik des Genialen. Mit Intuition, Kreativität und Intelligenz Probleme lösen*. Gabler Verlag, Wiesbaden, 1993.

- OSTERLOH, MARGIT (2012): Schluss mit dem Unfug. Sichern Rankings wirklich die Qualität der Universitäten? Eine Replik. *DIE ZEIT* 43, 18.10.12, p. 63.
- SCHEFOLD, BERTRAM & LENZ, THORSTEN [Hrsg.] (2008): *Europäische Wissensgesellschaft. Leitbild europäischer Forschungs- und Innovationspolitik?* Akademie-Verlag, Berlin, 2008.
- SCHMOCH, ULRICH: Messkonzepte für Innovationskultur. In: SCHEFOLD, BERTRAM & LENZ, THORSTEN [Hrsg.] (2008): *Europäische Wissensgesellschaft*, p. 99 – 111.
- SCHWEITZER, JAN (2014): Das hilft alles nichts. Sie halten sich schon sehr lange, obwohl längst erwiesen ist, dass sie nicht stimmen. Warum spielen viele vermeintliche Gesundheitsweisheiten trotzdem eine so große Rolle in unserem Leben? *DIE ZEIT WISSEN* 43, 16.10.14, p. 29 – 31.
- SCHWERTFEGER, BÄRBEL (2014): *Die Welt* 01.02.14.
- SPIEWAK, MARTIN (2005): Wächter der Forschung. *DIE ZEIT* 34, 18.08.05, p. 30.
- SPIEWAK, MARTIN (2005): Auf die Spitze getrieben. Der Elitewettbewerb verändert die deutschen Universitäten mehr als jedes Gesetz. Forscher und Politiker haben hochfliegende Pläne. *DIE ZEIT* 47, 17.11.05, p. 45.
- SPIEWAK, MARTIN [Gesprächsführer zwischen PLUMPE und ZIEGELE] (2010): »Sie führen Studenten in die Irre«. *DIE ZEIT* 16, 15.04.10, p. 69.
- SPIEWAK, MARTIN [Interview mit EINHÄUPL] (2011): »Das ist nicht seriös«. *DIE ZEIT* 35, 25.08.11, p. 63.
- SPIEWAK, MARTIN & WIARDA, JAN-MARTIN [Gesprächsführer zwischen MÜNCH, MÜLLER-BÖLING und HARTMANN] (2008): Wie nützlich sind Rankings? *DIE ZEIT* 22, 21.05.08, p. 65.
- STARKE, F. C. (1993): *Krise ohne Ende? Parteidemokratie vor neuen Herausforderungen*. Bund-Verlag, Köln, 1993
- STEGMÜLLER, WOLFGANG (1970): *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie Bd. II: Theorie und Erfahrung*. Springer-Verlag, Berlin, Heidelberg & New York, 1970.
- WIARDA, JAN-MARTIN (2009): Auf die Plätze, fertig – los! Europäische Universitäten schneiden schlecht ab in den weltweiten Rankings. Sie kritisieren deren Erhebungsmethode. Versprechen neue Verfahren mehr Gerechtigkeit? *DIE ZEIT* 36, 27.08.09, p. 69f.