

Prof. Fabio Melotti | AC  
Università di Padova  
Istituto Tricologico  
Sperimentale  
Corte Croce Valdeseo M

AdS,

F. METELLI

QUALITATIVES UND QUANTITATIVES IN DER EXPERIMENTELLPSYCHOLOGISCHEN FORSCHUNG

Meine Damen und Herren,

als mich Herr Professor Metzger zu diesem Vertrag einlud, dachte ich, dass ich am liebsten mit den Münsterschen Psycholegen einige Probleme besprechen würde, denen ich in meinen Forschungen begegnet bin, und denen ich nicht entgehen konnte, da ich ohne ihre Lösung nicht nur nicht ruhig weiterarbeiten konnte, aber sogar über den Sinn meiner Arbeit in Zweifel geriet.

Der Ausgangspunkt ist eine ganz gewöhnliche, und sogar banale Episode. Als ich vor zwei Jahren meinen berühmten Kollegen und Freund Sir F. Bartlett in Cambridge besuchte, zeigte ich ihm ein optisches Ergänzungssphänomen über das ich seit mehreren Jahren arbeitete; aber obwohl ich mich bemühte, die erforderlichen Bedingungen herzustellen, sah er das Phänomen nicht.

Es ist natürlich kein Wunder, dass eine gewisse Versuchsperson ein Wahrnehmungssphänomen nicht erlebt, und obwohl es unangenehm sein kann, dass eben der Wissenschaftler mit dem man das Phänomen zu besprechen beabsichtigt, es nicht wahrnimmt, doch ändert das an der Sache nichts. Die Lage wird aber kritischer, wenn eine Erscheinung, die den Kern einer Theorie bildet, von den Gegnern der Theorie nicht wahrgenommen und deshalb bezweifelt wird. Vielleicht war ich deshalb gegen die kleine Enttäuschung überempfindlich, weil ich vor kurzer Zeit einer ähnlichen Schwierigkeit begegnet war. Zusammen mit meinem Kollegen Prof. Kanizsa hatten wir eine Erscheinung demonstriert, dessen

Existenz Professor Michotte in seinem weltberühmten Buche über Ursachenwahrnehmung ausgeschlossen hatte, und die eine spezielle Seite der Michotteschen Theorie unhaltbar machte. Als wir aber die Sache Herrn Michotte mitteilten und ihm unsere Versuchsmittel zur Verfügung stellten, damit er selbst unsere Versuche wiederholen könnte, stellte sich für die meisten Versuche heraus, dass, was wir und unsere Mitarbeiter sahen, von Herrn Michotte und seinen Mitarbeitern nicht gesehen wurde.

Natürlich wurde die Schwierigkeit gelöst, indem die Versuche mit einer grossen Zahl naiver Versuchspersonen unter sehr strengen Versuchsbedingungen wiederholt wurden. Das ist aber, was ich eine juristische Lösung nennen würde. Mir kommt es vor, als ob man mit einer ~~Abstimmung~~ Wahlakt, zwischen zwei wissenschaftlichen Theorien entscheiden wollte. So kommt man zu einer abjektiven praktischen Entscheidung (nämlich welche der beiden Beschreibungen gelten soll); das wirkliche Problem wird aber nicht gelöst.

In der Wahrnehmungspychologie arbeitet man oft genug mit Erscheinungen, die so konstant über die individuellen Verschiedenheiten sind, dass es vorkommen kann, dass man solchen Schwierigkeiten überhaupt nicht begegnet. Doch habe ich immer bemerkt, dass man auch in diesen Fällen, wenn man nur mit einer grösseren Menge Versuchspersonen experimentiert, doch jemanden findet der, was alle anderen sehen, nicht sieht. Also besteht das Problem theoretisch immer.

Wenn man aber mit weniger konstanten Erscheinungen arbeitet, drängen sich die Deutungsprobleme nicht nur theoretisch sondern auch praktisch auf.

Wie soll man die Tatsache deuten, dass jemand eine gewisse Erscheinung nicht erlebt? Genügt es, dass eine einzige Versuchsperson eine Erscheinung wirklich unter kontrollierten Bedingungen erlebt,

damit - wie Jahannson meint - die Erscheinung als gesichert zu betrachten sei? Welche Bedeutung soll der verschiedenen Häufigkeit beige messen werden, mit der zwei verschiedene Erscheinungen erlebt werden?

Auf anderen Gebieten unserer Wissenschaft - besonders in der angewandten Psychologie - wo man quantitativ arbeitet, wird regelmässig mit individuellen Verschiedenheiten gearbeitet, deren Probleme mit Hilfe der Begriffe und der Technik der induktiven Statistik ohne besondere Schwierigkeit bewältigt werden. Wo nicht auf die Individuen sondern auf das Gemeinsame und Allgemeine eines Phänomens oder eines Prozesses gezielt wird, erreicht man ja die Parameter der Bevölkerung durch Schätzungsprozesse und Sicherheitsmasse. Kann man und soll man die selben oder ähnliche Methoden auch auf qualitative Gegebenheiten (wie die oben erwähnten) anwenden, und welche Bedeutung und Wichtigkeit soll ihnen beigemessen werden?

Die Lösung dieser Probleme hängt von dem Sinn des Gegensatzes Qualität-Quantität in der wissenschaftlichen Psychologie ab; ein grundsätzliches Problem, das man hier in keiner Weise vermeiden kann.

Quantitativ bedeutet zahlenmässig. Die offensichtliche Armut der quantitativen Data - die, wenn man sie mit der Fülle und Vitalität der phänomenal-qualitativen, trocken und so wie verstorben erscheinen - wird durch die Anwendbarkeit des übermächtigen Instrumentes der Mathematik kompensiert. Wie kommt es aber dazu, dass die Erörterungen, die Schlüsse, die für reine, ganz anschauungsfreie vernunftmässig definierte Begriffe gültig sind, auf Anschauungsdata verwendbar sind? Die Antwort liegt nahe: die Anwendbarkeit kann nur auf eine Ähnlichkeit begründet sein, die zwischen Anschauungsdaten und den Grundbegriffen der Mathematik, den Zahlen, besteht. Wenn eine solche Verwandtschaft besteht und inwiefern sie besteht, kann man die Gegebenheiten, die

wissenschaftlichen Beobachtungen und Resultate einfach wie Zahlen behandeln.

Dann kommt aber gleich für jede Wissenschaft und besonders für die wissenschaftliche Psychologie das wichtige und schwere Problem der Grenzen dieser Verwandtschaft, und folglich der korrekten Anwendung der Mathematik.

Hier kann ich mich an die wohlbekannte Analyse Stevens stützen. Stevens unterscheidet vier Messungsmodelle, die sich schrittweise dem mathematischen Modell nähern, und die <sup>schrittweise</sup> fortlaufend neue mathematische Operationen zulassen: in der erweiterten Bedeutung deckt sich der Begriff "Messung" mit dem der "quantitativen Auffassung" vollkommen. (Fig. 1)

Obwohl es sich um Bekanntes handelt, erörtere ich kurz die Stevenschen Modelle, die ~~z~~ den Ausgangspunkt unserer Problematik bilden. Das letzte Modell, ist durch den vollkommensten Isomorphismus mit der Zahl charakterisiert; es hat sein Grundgebiet in der Physik, und ist in der Psychologie kaum repräsentiert, und interessiert uns deshalb weniger als die drei ersten. Das nächstfolgende ist durch die Gleichheit der Intervalle zwischen sukzessiven Stufen, oder, in anderen Worten, durch die Konstanz der Messungseinheit, aber auch durch die Willkürlichkeit des Nullpunktes charakterisiert. Ein typischer Beispiel einer solchen Messungsskala bilden die gewöhnlichen Temperaturskalen (Celsius, Fahrenheit); sämtliche Messungen, die man in der Testpsychologie mit Standard-Einheiten ausführt, sollen zu diesem Modell gehören.

Für unser Problem sind aber die zwei niedrigeren Formen der Messung wichtiger. In einem Modell entbehrt die Messung nicht nur des absoluten Nullpunktes, sondern auch der Gleichheit der Intervalle.

Die Gegebenheiten bilden eine monotone Folge, man kann mit ihnen

Fig. 1

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)	Typical Examples
Nominal	Determination of equality	Permutation group $x' = f(x)$ ( $f(x)$ means any one-to-one substitution)	Number of cases Mode Contingency correlation	"Numbering" of football players Assignment of type or model numbers to classes
Ordinal	Determination of greater or less	Isotonic group $x' = f(x)$ ( $f(x)$ means any increasing monotonic function)	Median Percentiles Order correlation (type 0)	Hardness of minerals Quality of, leather, lumber, wool, etc. Pleasantness of odors
Interval	Determination of equality of intervals or differences	General linear group $x' = ax + b$	Mean Standard deviation Order correlation (type I) Product-moment correlation	Temperature (Fahrenheit and centigrade) Energy Calendar dates "Standard scores" on achievement tests (?)
Ratio	Determination of equality of ratios	Similarity group $x' = ax$	Geometric mean Coefficient of variation Decibel transformations	Length, weight, density, resistance, etc. Pitch scale (mels) Loudness scale (sones)

eine Rangfolge bilden, der Abstand zwischen den nächstfolgenden Stufen ist aber nicht gegeben. Im niedrigsten Messungsmodell besteht auch keine Rangfolge, es gibt nur eine Kategorisierung, eine Klasseneinteilung; jede Klasse enthält einen oder mehrere Klassenglieder, die als Klassenglieder gleich sind.

Betrachten wir näher die zwei niedrigeren Modelle, so sehen wir, dass sie das Gebiet der sogenannten qualitativen Daten vollständig decken. Was das letzte Modell betrifft, ist ja nicht unsere Sprache, unser naives übliches sprachliches Denken im Grunde eine ständige Kategorisierung? Denn nicht nur die wissenschaftliche Begriffsbildung, sondern auch jede sprachliche Schilderung unterscheidet sich vom anschaulich Unmittelbaren und als solches Unaussprechbaren durch eine ständige Klasseneinteilung.

Ich will aber nicht näher auf die übliche Folge der Stevenschen Schilderung, dass wir alle immer und notwendigerweise messen, eingehen, und beschränke mich auf die für meinen Gedankengang wichtige Feststellung: dass es keine Kluft zwischen sogenannten qualitativen und quantitativen Gegebenheiten besteht, dass der Unterschied eher des Grades als des Wesens ist.

Auf jedem Niveau sind also gewisse mathematische Operationen erlaubt, gewisse quantitative Methoden korrekt anwendbar. Wir wollen vorläufig die Frage nicht erörtern, inwiefern die Tatsache, dass gewisse quantitative Methoden anwendbar sind, uns zu ihrer Anwendung zwingt, und gleich zum wichtigen Problem der Anwendung der Stevenschen Modellen übergehen.

Aus welchen Gründen soll es festgestellt werden, welches Modell im konkreten Fall anzuwenden sei? Ist es möglich (und notwendig) auf jedem Gebiet, oder sogar in jedem Fall objektiv festzustellen, welchem Modell die Tatsachen gleichen, oder ist es eine Sache der Zweck-

*bestätiges*

mässigkeit, oder sogar der Willkür, ein gewisses Modell zu wählen?

Das Problem ist nicht leicht zu lösen. In unserer Wissenschaft sieht man, dass die Mathematisierung auf verschiedenen Gebieten verschieden ist, unabhängig von dem Entwicklungsgrade und dem erreichten Niveau der Forschungen auf jedem Gebiet, was für eine spezifische Beziehung zwischen Gebiet und Modell spricht. Anderseits gibt es Gründe für die Willkürlichkeit der Anwendung der quantitativen Methodik. Nicht nur, dass verschiedene Forschungen auf gleiche oder sehr ähnliche Data verschiedene Messungsmodelle anwenden, sondern verschiedene Forscherschulen und Gruppen gegenseitige Meinungen über das Wesen der Psychologie verfechten; wenn man nur die extremen <sup>Teilige</sup> Lagen betrachtet, gilt für die einen die Gleichung Wissenschaft=Quantitativ, und das Nicht-quantitative (im Sinne der zwei niedrigeren Stevenschen Modellen) ist keine Wissenschaft, sondern Vorwissenschaft; während für die anderen nur das Individuell-Einzige als echt, jede Generalisierung in der Psychologie als oberflächlich, grob, unwesentlich, unecht zu betrachten ist.

Eine Lösung kann man vielleicht doch finden durch die Konstaterung, dass die Wissenschaft immer durch die Berücksichtigung des - für ein gewisses Problem - Wesentlichen und die Vernachlässigung des Unwesentlichen fortgeschritten ist. Das individuell+Einzig ist das Objekt der Kunst, nicht der Wissenschaft. Galilei hat ja von dem Unwesentlichen, dass das Fallen eines Blattes und eines Steines individuell-verschieden mache, absehen müssen, um das Allgemeinwesentliche heraus zuarbeiten. Es ist auch klar dass man dann, wenn die so gewonnenen Resultate einem höheren Messungs-Modell gleichen, auf die Anwendung des mächtvollen Instrumentes der Mathematik nicht verzichten kann.

Nur scheint mir, die psychologische Forschung zu lehren, dass

die beiden Forderungen - Abstrahierung des Wesentlichen und Anwendung eines Messungsmodells höheren Grades - oft nicht zusammen passen.

Ein Beispiel aus meiner Forschungsarbeit wird vielleicht den gemeinten Sachverhalt näher erklären.

In der klassischen Aussageforschung wurden oft genug, seit den Sternschen Bilderbeschreibungsexperimenten, die Aussagefehler studiert. Das Bild, das man den Versuchspersonen zeigen soll, oder der kurze Film, den man ihnen vorführen soll, wird nach einer möglichst vollkommenen Beschreibung in Elemente zerstückelt, so dass man durch einen Elementarvergleich, die Zahl der richtigen, und eventuell auch der falschen und der unterlassenen Elemente jeder Aussage ermitteln kann. Das so bearbeitete Material funktioniert dann wie ein Instrument, mit dem man Ermittlungen z.B. über die verschiedene Aussageeignung verschiedener Menschenklassen (Männer, Frauen, Jugendliche, Erwachsene) unter verschiedenen Bedingungen (Ermüdung, Emotion u.s.w.) macht.

Man wird mit Vorteil das Experiment nach einem der Fischerschen Experiment-Plänen entwerfen und so das kostbare Instrument der Varianzanalyse anwenden können.

Sehen wir uns nun die angewandte Methode näher an. Die richtig reproduzierten Elemente gelten als Indices der Leistungshöhe. Hier wird also (im (gewöhnlichen Sinn) gemessen, und die vollzogenen Operationen sagen uns, dass man ein der zwei höheren Stevenschen Modelle angewandt hat: die Stufen werden als gleich behandelt, doch, ist kein absolutes Nullpunkt erforderlich; so können wir uns hier auf das dritte Modell beziehen.

Soll die Anwendung des Modells als korrekt gelten? Man kann die Behauptung nicht ohne Verbehalt annehmen. Nicht dass man prinzipiell

nicht berechtigt sein sollte, die bunte Verschiedenheit der Beschreibungen in eine einzige Klasse zu fassen; es bestehen aber ernste Zweifel über die Berechtigtheit, eine konkrete Beschreibung nach einem vorgebildeten Schema zu zerstückeln, und noch mehr darüber, dass die Stücke als gleich zu betrachten seien.

Es ist ja von vornherein klar, dass die sogenannten Elemente in diesem Fall nicht gleich sein können, auch wenn man vom praktischen Gesichtspunkt der Leistung ausgeht. Man kann aber antworten, dass die Tatsachen nie den mathematischen Modellen gleichen können, dass es sich immer um mehr oder weniger grosse Annäherungen handelt. Die Entscheidung über die Anwendbarkeit des Modells kann man nur durch indirekte Kriteria treffen. In diesem Fall sollte man z.B. direkt aus den Experimenten ausgehend, eine gewisse Kontrolle ausüben: wenn vernünftige Erwartungen nicht erfüllt werden (wenn z.B. die frischen Versuchspersonen nicht besser als die ermüdeten aussagen) hat man schon Indizien - die man nach Willkür vermehren kann - dass die Anwendung des Modells unberechtigt ist.

Doch kann man das Problem auch von einer anderen Seite behandeln. Als ich meine ersten Schritte in der wissenschaftlichen Psychologie machte, arbeitete mein verehrter Lehrer auf dem Gebiet der Aussagepsychologie, eben mit einer ziemlich komplizierten quantitativen Methodik, die sich auf die oben erwähnte Fragmentierung der individuellen Aussageprotokolle fundierte. Er studierte eigentlich das interessante Problem der optimalen Latenzzeit, da es sich paradoxer Weise herausgestellt hatte, dass unter gewissen Umständen die nicht unmittelbaren Aussagen weniger fehlerhaft waren als die unmittelbaren.

Als ich aber die Protokolle las, fiel es mir ein, dass man unter dem Begriff "Fehler" eine unglückselige Kategorisierung mache, da

als Fehler etwas Negatives verstanden wird, während es sich um positive Leistungen handelte. Wenn man dieselben Tatsachen anstatt als Fehler, als Verwandlungen, als Umgestaltungen betrachtete, war ihr Sinn gleich zu begreifen; und man sah auch sofort, dass dieser Sinn nur ausnahmsweise aus der elementarischen Einheit, die aus Messungszwecken gebildet worden war, zu verstehen war, sondern in der Regel mehr oder weniger breite Gebiete der Filmhandlung anging.

Die Aussagefehler konnten also von einem ganz verschiedenem Standpunkt studiert werden, und zwar als positive Leistungen anstatt als "nicht gelungene Reproduktionen".

Welches sind die Vorteile und die Nachteile der beiden Wege, die sich zur Erforschung desselben Materials anbieten?

Es ist nicht schwer zu zeigen, dass der zweite Weg, die Erforschung der Fehler als Umgestaltungen zu einer tieferen Einsicht bringt. Man bemerkt, dass sich die Fehler zum Filminhalt, zur Episode, die den Versuchspersonen vorgeführt wurde, in einer charakteristischen Weise gruppieren. Es gibt Stellen, die einer Änderung weniger widerstehen. Noch haben, im Allgemeinen die Änderungen an einer Stelle dieselbe Richtung, und unterscheiden sich nur an der Intensität. Der erlebte Inhalt scheint gewissen Spannungen ausgesetzt zu werden, die gewisse Stellen umzuwandeln tendieren, und denen ein individuell verscheidener Widerstand geleistet wird. Andere Umformungen sind nicht auf gewisse Stellen beschränkt, sondern betreffen im Allgemeinen den ganzen Inhalt, der tief umgewandelt wird; in dem Fall wird der Sinn anders erlebt, und dem erlebten Sinn schmiegen sich die Einzelheiten an.

Die zu-letzt geschilderten, zentralen Unwandlungen sind oft individuell verschieden, doch ist ihre Verschiedenheit beschränkt, und besonders aus den verschiedenen Auffassungsmöglichkeiten des

Ereignisses zu verstehen.

Wenn im Vergleich mit den hier kurz geschilderten Deutungsmöglichkeiten, die Oberflächlichkeit der Einteilung "richtige Reproduktionen" - "Fehler" und "Unterlassungen", und die Unmöglichkeit mit dieser Methodik irgend eine Einsicht in die Gestaltungs dynamik der Aussagen zu gewinnen, klar werden, so muss man doch auch zugeben, dass die verfeinerte <sup>statistische</sup> Fischersche Methodik nur bei der einfachen, elementaristischen Einteilung anwendbar ist.

Und hier kommen wir endlich auf das allgemeine methodische Problem zurück, das wir an einem konkreten Fall erleuchtet wollten. Wenn die zwei Forderungen - Abstrahierung ~~der~~ des Wesentlichen und Anwendung eines Messungsmodells höheren Grades - unvereinbar sind, wie soll die Wahl getroffen werden? Und, ist das Messungsmodell willkürlich zu wählen, oder sachlich bestimmt?

Er scheint mir, dass man die beiden Fragen zugleich beantworten kann.

Sind an jedem Tatsachenmaterial mehrere Messungsmodelle anwendbar? Wir haben an einem konkreten Fall gesehen, dass man zwei weit verschiedene Modelle anwenden konnte, ein verfeinert quantitatives und ein grob-kategoriales; die Antwort scheint also eine bejahende zu sein. Doch haben wir gesehen, dass nicht nur die Resultate weit verschieden waren, sondern auch die Tatsachen, auf die sich die Resultate bezogen - und zwar die Leistungshöhe einerseits, die Gestaltungsprozesse anderseits verschieden waren. So scheint es, dass es doch nur ein geeignetes Modell für jedes Forschungsproblem gibt; verschiedene Modelle bringen mit sich verschiedene Umformungen des Materials und ändern notwendigerweise die Probleme. Es folgt auch das die Problemstellung oder, was im Grunde gleich ist, die Abstraktion des Wesentlichen das Wichtigste ist, nicht das Messungsmodell, das Niveau der Mathematisierung; wo die Forderung der quantitativen

Bearbeitung im Vordergrund steht, kommt es zu dem gut bekannten Gegensatz zwischen dem Luxus der mathematischen Methodik und der Dürftigkeit der Resultate.

2. Durch die vorhergehende Analyse sind wir zum Schluss gekommen, dass gewisse Probleme eine niedere Stufe der Abstraktion fordern, der das niedrigste Messungsmodell entspricht: sonst wird eben das Wesentliche, das man abstrahieren sollte, vernachlässigt, und wie man sagt, das Kind mit dem Bade ausschüttet.

Da diese Lage wahrscheinlich oft in der experimentellen Psychologie zu finden ist, scheint es mir angemessen, die Beziehungen der bekanntesten und wichtigsten quantitativen Methoden zum kategorialen Messungsmodell näher zu untersuchen. Da wir nun gezeigt haben, dass die ungünstigen Folgen der Mathematisierung nicht von der mathematischen Methodik, sondern von der ungünstigen Wahl des Abstraktionsniveaus, und des <sup>zugehörigen</sup> übereinstimmenden Messungsmodells abhängt, besteht kein Grund mehr, die Möglichkeiten des richtigen Messungsmodells nicht völlig auszubeuten.

Die Varianzanalyse ist, wie üblich, ein Mittel um hauptsächlich:

- die Wirkung mehrerer Bedingungen gleichzeitig zu ermitteln;
- die Wirkung einer Bedingung zu beseitigen, um andere Bedingungen hervorzuheben;
- zu bestimmen, ob die Wirkung zweier oder mehrerer <sup>zusammen</sup> ~~nebenwirkenden~~ Bedingungen verschieden von der Summe ihrer isolierten Wirkungen ist.

Diese ganz Kurze, und unvollkommene Schilderung genügt schon, um einen Eindruck der Wichtigkeit der Methode, zu bekommen.

Doch wird die Methode [nicht so oft gebraucht, wie man es erwarten würde, und, was sonderbar ist, wird sie] oft an uninteressante Probleme angewandt, die sicher den Aufwand an Methodik nicht rechtfertigen, X

X oder; zugleich

und manchmal sogar den Eindruck geben, nur den Anlass zu sein, eine komplizierte Methode anzuwenden, wie gewisse Forschungen die den Verdacht auferwecken, dass sie nur, um eine gewisse Apparatur zu <sup>be</sup>brauchen, geplant wurden. ~~nutzen~~

Es ist klar, dass die letzterwähnten Fälle hauptsächlich aus menschlichen Gründen zu verstehen sind. Doch hängt die oben erwähnte Lage von tatsächlichen Schwierigkeiten ab, die der Methode innenwohnend sind, und zwar:

1. Die Notwendigkeit von ~~Vornherein~~ einen starren Experimentierungsplan zu entwerfen, an den man sich halten muss, anstatt nach Versuchen fortzuschreiten, wie es in der echten Forschung gewöhnlich vorkommt.
2. Die Notwendigkeit, sich an gewissen fixen Experimentierungsschemen ~~anzuhalten~~, die die nicht-zu-unterschätzenden Vorteile einer tadellosen statistisch-mathematischen Entwicklung anbieten, deren aber eine beschränkte Zahl besteht, und die nur relativ den speziellen Forderungen einer Forschung anpassungsfähig sind.
3. Die Notwendigkeit, dass die zu bearbeitenden experimentellen Resultate wenigstens dem 3. Stevenschen Messungsniveau gehören, also quantitativ, im üblichen Sinn des Wortes, seien.

Es sind, offenbar, besonders schwere Einschränkungen, hauptsächlich die Letzte, die sämtliche Untersuchungen ausschliesst, deren Resultate extrem verarmt oder sogar des Sinnes beraubt würden, wenn man sie in Messungseinheiten ausdrücken wollte.

Doch sind die Vorteile der Methode so gross, dass man sich fragen muss, ob es wirklich keine Mittel gibt, um die genannten Schwierigkeiten zu überbrücken.

Wir werden an einem konkreten Problemkreis sehen, ob und wie die genannten Schwierigkeiten zu überwinden sind. Früher aber wollen wir auch die Faktorenanalytische Methodik kurz besprechen.

Wie bekannt, ist die Faktorielle Analyse eine Methode die haupt sächlich aus einer grüsseren Zahl "willkürlicher Veränderlichen" aus gehend, die Grundveränderlichen zu ermitteln erlaubt.

Die so gewonnenen Resultate hängen offenbar von der Problemstellung ab, und können die Daten, die im Problem gebraucht wurden, nicht überschreiten. Die Methode kann in einem ganz ungeklärten Problem gebraucht werden, um eine vorläufige Klärung zu erzielen: in diesem Fall soll das Resultat, das nicht einer gewissen Willkürlichkeit entbehren kann, nicht als eine endgültige Lösung, sondern als eine Hypothese, die mit anderen Mitteln kontrolliert werden soll, angenommen werden. Es ist aber vielleicht weniger bekannt, dass die Methode auch für einen anderen Zweck gebraucht werden kann, und zwar zur Kontrolle einer ganz bestimmten Hypothese; die Faktorenanalytische Methode kann, wie allerdings jede statistische Methode, nicht eine Hypothese direkt beweisen; sie kann aber die Gründe hervorheben, weshalb eine Hypothese zu verwerfen ist, oder die Aussage berechtfügen, dass es Gründe gibt, die zugunsten der Hypothese sprechen.

Die Erfindung der Faktorenanalytischen Methode ist ein Ruhm der Testpsychologie, und es ist deshalb kein Wunder, wenn die Methode hauptsächlich zur Lösung von Eignungsproblemen und zur Erforschung von Persönlichkeitsstrukturproblemen, durch Tests und andere Schätzmethoden, gebraucht worden ist. Doch scheint die Methode prinzipiell nicht auf diese und ähnliche Problemkreise beschränkt zu sein.

Soll man also zum Schluss kommen, dass die beschränkte Anwendung der Methode nur in der wissenschaftlichen Tradition ihre Gründe hat?

Wenn man die Methode einer näheren Untersuchung unterwirft, treten Vorteile und Nachteile hervor. Ein grosser Vorteil der Methode ist, dass sie auch auf nicht quantitativen Daten (im üblichen Sinn) anwendbar ist: sie ist auf die Korrelationsrechnung fundiert, und Korrelationen kann man auch mit einfachen Klassifizierungen des

dychotomischen Typus (z.B. Bewegung-Ruhe, Körperlich-Flach, u.s.w.) rechnen. Es gibt natürlich auch Nachteile. Aber die scheinbar grösste Beschränkung besteht in der Notwendigkeit mit einer grossen Zahl Ver suchspersonen zu arbeiten. Für eine Analyse, die auf nur 10 Ausgangs veränderlichen beschränkt ist, sind mindestens 200-300 Versuchspersonen notwendig. Wenn man, wie oft, individuell experimentieren muss, und die Experimente Komplex sind und eine gewisse Eindübung voraussetzen, dann bildet die oben erwähnte Forderung ein unüberwindliches Hindernis.

3. Die kritische Betrachtung einiger der wichtigsten modernen quantitativen Methoden hatte den Zweck an einem konkreten Forschungsgebiet die Probleme des Gegensatzes Qualität-Quantität womöglich zu demonstrieren. Als Forschungsbereich habe ich das Gebiet der optischen Wahrnehmung gewählt. Dazu hat mich besonders ein Grund geführt: das Wahrnehmungsgebiet ist ein solches, wo die fruchtbarsten Untersuchungen sich im phänomenal-qualitativen entwickelt haben, wo die quantitativen Methoden von Anfang an zu dürftigen Resultaten verurteilt zu sein scheinen. Ich glaube aber, dass hier keine prinzipielle Schwierigkeit bestehe, dass man auch in diesem wichtigen Gebiet die quantitative Methodik mit Vorteil brauchen kann, wenn man sie nur am richtigen Platz braucht, und dem Grundsatz getreu bleibt, keine Gewalt dem Phänomenalen zu tun. Die quantitative Methodik soll sich der speziellen Beschaffenheit der Daten anpassen, nicht die Daten der Methodik.

Sehen wir uns die Forschungsarbeit im Wahrnehmungsgebiet näher an. Wie bekannt, besteht in der gesamten Psychologie, aber vor allem in der Wahrnehmung eine besondere Schwierigkeit, die Probleme zu sehen. Ein Grund davon ist die natürliche Einstellung, die uns so wirksam Harr Professor Metzger in seinen Gesetzen des Sehens geschildert

hat: "man öffnet die Augen und die Welt ist da". Deshalb ist es so schwer zu schildern (nicht nur den Laien, sondern auch manchen Psychologen) dass in der Wahrnehmung alles zu erklären ist, nicht nur die optischen Täuschungen.

Die erste Merkwürdigkeit der Wahrnehmungsforschung ist also, dass man gewöhnlich nicht mit einer Eßungsfrage sondern mit einer Problemfrage anfängt. Und die Entdeckung des Problems ist oft nicht weniger schwer als die Entdeckung der Lösung.

Der Anfang ist also ganz aussergewöhnlich und die Fortsetzung ist nicht weniger revolutionär, wenn man sie vom Standpunkt der traditionellen Methodik betrachtet.

Ich wüsste sie auch nicht nach allgemeinen Richtlinien zu beschreiben. Es ist viel leichter zu sagen was auf dieser Forschungsstufe nicht geschieht, als was eigentlich geschieht. Sicher denkt kein Mensch an unabhängige und abhängige Veränderlichen, oder an die Einführung statistischer Kontrolle, noch an geplantes Experimentieren. Gegen jeden methodologischen Grundsatz, ist der Versuchsleiter auch Versuchsperson, wenn auch mit der gelegentlichen Kontrolle einiger Mitarbeiter; denn eine wesentliche Bedingung dieser Forschungsstufe ist, möglichst schnell von einem Versuche zum anderen fortzuschreiten, um die Richtungen in denen die Forschung fruchtbar ist, zu entdecken. So kommt man, durch sukzessive Versuche, zu mehreren kleinen Entdeckungen, (in Bezug auf die Weisen, in denen sich das Phänomen unter verschiedenen Bedingungen gestaltet); bis man es zuletzt, wenn möglich, fertig bringt, die Erscheinung von allen akzessorischen Bedingungen zu befreien, und sie nur mit den (wesentlichen), notwendigen Bedingungen zu reproduzieren.

Ohne Zweifel ist das der schönste, der faszinierendste, aber auch der schwierigste Teil der Forschung, den man weder direkt noch aus den Büchern lernen kann, da er nur von den persönlichen Gaben

des Forschers abhängt. (Wenn man denkt, dass oft ein junger Student der philosophischen Fakultät, sich dem Psychologischen Laboratorium nähert um "die wissenschaftliche Methode" zu lernen, soll er doch einen groben Chock erleiden, wenn er einsieht, dass eben in der wesentlichsten Phase der Forschung keine echte Methode besteht, die zu befolgen ist).

Die Forschung kann auch schon auf dieser Stufe zum Abschluss kommen, wenn nur eine allgemeine Kontrolle der Resultate auf einer auch beschränkten Zahl von Versuchspersonen vorgenommen worden ist, die dazu genügen, dem Forscher die Objektivität seiner Beobachtungen zu sichern.

Damit kann entweder die beschreibende Phase der Arbeit abgeschlossen werden - und so ist bestimmt genug geleistet worden, wenn neue Tatsachen zur Schau getragen wurden - oder die gewonnenen Resultate gestalten sich derart, dass sie eine Erklärung oder sogar eine Theorie zu gründen erlauben. Man muss aber zugeben, dass somit nur die erste - wenn auch die schönste, die schwierigste, die wichtigste - Phase der Untersuchung abgeschlossen worden ist.

Bevor ich zur zweiten Phase übergehe, ist vielleicht eine Erklärung am Platz. Diese Phasenteilung, ist wie jede Teilung eines natürlichen Prozesses, künstlich, sie ist nicht immer anwendbar - man kann sogar mit Richtigkeit behaupten, dass die beschreibende Phase nie vollständig abgeschlossen ist, und deshalb notwendigerweise auch während der nächsten Phase fortschreitet; und dass in der ersten Phase oft Bearbeitungen zu finden sind, die eigentlich der nächst zu schilddernden zweiten Phase gehören.

Die zweite Phase wird oft durch andere Forscher entwickelt, und deshalb sind oft die betreffenden Untersuchungen durch theoretische Gegensätze, Kritik, Berichtigungen und Polemik charakterisiert. Doch ist diese Seite nebensächlich: natürlich wird, wenn auch kein Gegen

satz der Gesichtspunkte besteht, in dieser Phase eine weitere Kontrolle ausgeübt, es werden die früheren Resultate präzisiert, vertieft, erweitert und oft teilweise berichtigt, aber das wesentliche ist, dass man in einem schon teilweise geklärten Versuchsgebiet, die statistischen Kontrolle einführen und eine oder mehrere Versuchsserien mit Vorteil planen kann. Es hat nämlich einen Sinn, die Forschung zu planen, nur wenn das Forschunggebiet soweit untersucht worden ist, dass man schon genügend fundierte Vermutungen über die Resultate machen kann.

Ich denke, dass es hier angemessen ist einen konkreten Anhaltspunkt zu suchen, um eine unnötige Schwierigkeit zu vermeiden. Von nun an werde ich mich auf die berühmte Untersuchung Michotte's über die Ursachenwahrnehmung beziehen. Das Buch Michottes ist, meiner Meinung nach, ein vorbildliches Beispiel der ersten Phase einer experimentellen Untersuchung; einige Beiträge die der zweiten Phase gehören, bestehen schon.

Die Resultate der Michotteschen Untersuchung sind im üblichen Sinn, qualitativ und entsprechen dem niedrigsten Messungsniveau Stevens. Die Kategorisierung ist, im Falle des "Stossens" zweifach (kausale Beziehung - <sup>oder</sup> Unabhängigkeit der Bewegungen) oder mehrfach (kausale Beziehung <sup>oder</sup> Auslösungsphänomen <sup>oder</sup> Unabhängigkeit), aber doch ohne Gewalt in zwei Klassen anordnungsfähig; im Falle des Schiebens - Ziehens, mehrfach (A bewegt B - B bewegt A - A und B bewegen sich unabhängig). Auf anderen Gebieten der optischen Wahrnehmung finden wir ähnliche natürliche Klasseneinteilungen der Phänomene, und zwar: Ruhe-Bewegung; Vereinigung nach wagerechten oder senkrechten Reihen; optimale Bewegung-duale Bewegung-reines  $\gamma$  - Ruhe u.s.w. In manchen Fällen handelt es sich um die Alternative Anwesenheit-Abwesenheit einer Erscheinung, in anderen Fällen, um verschiedene Erscheinungen, die sich durch Varierung der (äußeren oder inneren)

Bedingungen vergegenwärtigen. In den meisten Fällen sind die Übergänge fließend (man unterscheidet z.B. mehrere Grade des Ursächlichkeitseindrucks, von einem ganz deutlichen, zwingendem Eindruck, bis zu einem blassen unsicheren Eindruck, der stufenweise den Unabhängigkeitseindruck erreicht, so dass man den Grad, oder wenigstens die Rangfolge zweier Eindrücke angeben könnte; in anderen Fällen (Typus Figur-Grund) gibt es vielleicht reine Alternative.

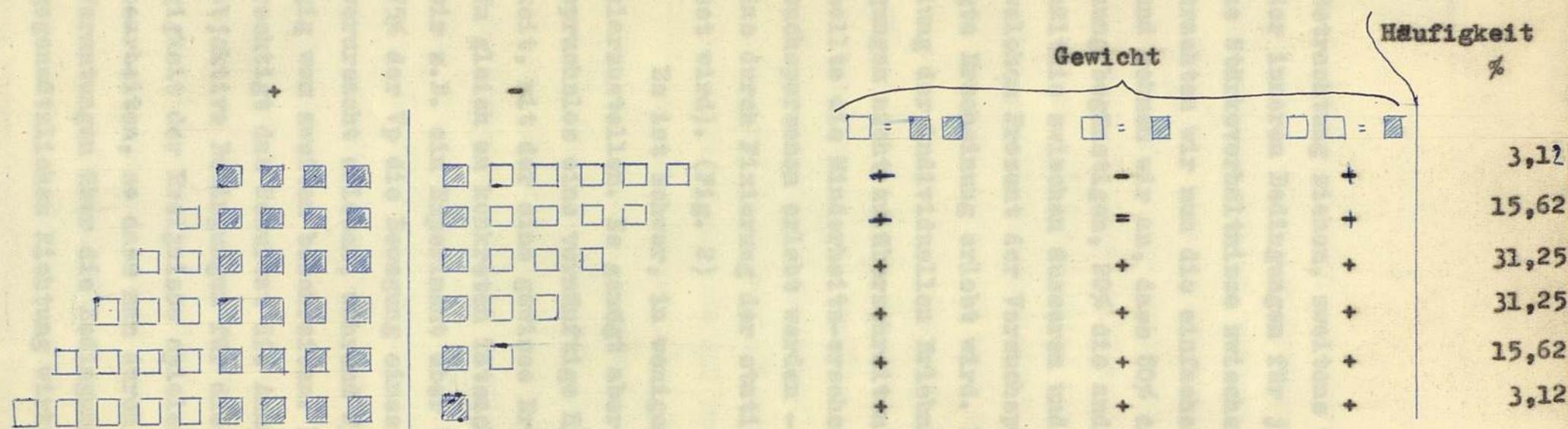
Es ist üblich, dass man an so gearteten Resultaten statistische Kontrollmethoden anwenden kann; es erhebt sich aber die Frage nach dem Sinn solcher Kontrolle;

Wir sind, nach einer langen Wanderung, wieder auf die Frage gekommen, die wir am Anfang gestellt hatten: 30% Vp erleben eine Erscheinung unter gewissen Bedingungen; wenn eine Bedingung variiert wird, wird die Erscheinung von 60% Vp erlebt. Das statistische Test sagt uns dass es höchst unwahrscheinlich ist, dass der Unterschied zufällig ist. Aber was der Unterschied bedeutet, erklärt die statistische Probe nicht. Handelt es sich vielleicht um eine verschiedene Volkstümlichkeit beider Phänomene? Was hat eigentlich die Zahl der Versuchspersonen, die es erleben, mit dem Phänomen zu tun? Welche Bedeutung soll, z.B., der Tatsache beigemessen werden, dass nur eine Minderheit ein gewisses Phänomen erlebt?

Ich glaube, dass man die Tatsachen folgenderweise erklären kann. Die psychophysischen Wahrnehmungsprozesse hängen von der Wirkung äusserer und innerer Bedingungen ab. Die äusseren, Reizbedingungen sind in erster Annäherung im selben Versuche für alle Versuchspersonen gleich, während die inneren Bedingungen, die man nur einer sehr beschränkten Kontrolle unterwerfen kann, für jede Vp. verschieden sein können.

Es ergibt sich ein Modell, das den Tatsachen zu passen scheint. Man soll erstens verschiedene Verteilungen der Reizbedingungen in

Fig. 2



Deutungsmodell einer dichotomischen Wahrnehmungserscheinung: unter den selben äusseren Bedingungen ereignen sich 2 verschiedene Wahrnehmungen (z.B. Pokal - Profil, in der bekannten Rubin'schen Zeichnung).

In der Abbildung werden 5 äussere und 5 innere Bedingungen gleichen Gewichts in Betracht gezogen; 4 äussere Bedingungen spielen für eine Alternative (+), eine für die andere Alternative (-).

Die inneren Bedingungen verteilen sich nach der zusammengesetzten Wahrscheinlichkeit.

- = Objektive Bedingungen (Reizbedingungen)
- = Subjektive Bedingungen (Innere Bedingungen: Einstellung, u.s.w.)
- =   Die objektiven Bedingungen wirken das doppelte der subjektiven Bedingungen

Betrachtung ziehen, zweitens verschiedene individuelle Verteilungen der inneren Bedingungen für jede Reizverteilung, drittens verschiedene Stärkeverhältnisse zwischen äusseren und inneren Bedingungen. Betrachten wir nun die einfache Alternative, zweier Erscheinungen und nehmen wir an, dass 80% der Reizbedingungen, die eine Erscheinung begünstigen, 20% die andere, dann hängt es von dem Stärkeverhältnis zwischen äusseren und inneren Bedingungen ab, ob und im welchem Prozent der Versuchspersonen die objektiv weniger begünstigte Erscheinung erlebt wird. Doch sieht man gleich dass die Verteilung der individuellen Erlebnisse die Verhältnisse der Reizbedingungen nicht zu überschreiten tendiert. (Also, in unserem Beispiele sollte die Minderheits-erscheinung von nicht mehr als 20% der Versuchspersonen erlebt werden - abgesehen von dem zufälligen Schwenken das durch Fixierung der statistischen Sicherheitsgrenzen ausgeschaltet wird). (Fig. 2)

Es ist schwer, in wenigen Worten das Funktionieren des Modells klarzustellen. Es genügt aber, zu zeigen, dass man anscheinend widersprüchliches eine vernünftige Hypothese über die Bedeutung der Häufigkeit, mit der eine gewisse Erscheinung erlebt wird, aufstellen kann. Um gleich an konkreten Tatsachen unsere Hypothese anzuwenden, nehmen wir z.B. ein Experiment über Ursachenwahrnehmung in Ansicht, in dem 75% der Vp die Bewegung eines Objektes als durch ein anderes Objekt verursacht erleben, während 25% die Bewegung des ersten als unabhängig von zweiten beschreiben. Nach den statistischen Kontrollen, berechtigt das Resultat die Annahme, dass in dem Fall bedeutend mehr objektive Bedingungen für die kausale Beziehung als für die Unabhängigkeit der Ereignisse spielen. Danach kann man die Annahme weiter bearbeiten, so dass man durch eine Analyse der Bedingungen konkrete Vermutungen über die Bedingungen, die in einer Richtung und in der gegensätzlichen Richtung wirken, stellen kann. Wenn man über die

Häufigkeitsbeziehung mehrerer Versuche verfügt, wird die Annahme et was fester; dieselben Bedingungen die man, in verschiedenen Zusammenhängen wieder findet, zwingen zu einer gemeinsamen, konsequenter Deutung, die weniger Bewegungsfreiheit zulässt, und deshalb die Annahme einigermassen kontrolliert.

Man kann aber vielleicht noch einen Schritt weiter gehen. Soll man durch den Unterschied der Häufigkeitprozente die Wirkung der verschiedenen Bedingungen nicht direkt beweisen können?

Es gibt natürlich auch einen altbekannten Weg, der uns dazu führt, wenn es gelingt nur eine einzige äussere Bedingung zu variieren, während die anderen Konstant bleiben. Wenn zwischen den beiden Fällen ein Unterschied der Häufigkeitprozente besteht, und der Unterschied statistisch gesichert ist, dann kann man die Wirkung der betreffenden Bedingung als bewiesen betrachten. Doch ist die Brauchbarkeit dieser Methode auf spezielle Fälle beschränkt.

Hier scheint aber eine viel wichtigere und mächtigere Methode brauchbar zu sein. Bisher haben wir sämtliche Versuchsresultate, auch diejenigen, die sich dichotomisch alternativ einteilten, als Klassen einteilungen bezeichnet, auf die man nur gewisse statistische Methoden anwenden durfte; eine Alternative entspricht aber einer Schätzungsskala, die - im allgemeinen nur um die Aufgabe der Vp zu erleichtern - über nur zwei Stufen verfügt, 1 und 0. Und auch eine solche grobe Messungsskala erlaubt, Mittel - und Strengswerte zu rechnen.

Sind wir also entgleist, und haben wir doch den qualitativen Daten ein Messungssystem aufgepresst? Ich glaube es nicht. Wir haben nichts an den Resultaten geändert, nicht Verschiedenes willkürlich für Messungszwecke zusammengehäuft. Die Resultate sind unverändert geblieben; nur haben wir konstatiert, dass man sie quantitativ deuten kann. Das logische Problem, wie wir vom ersten zum dritten Steven'schen Niveau übergegangen sind, muss dahingestellt blei-

ben. Wenn die obigen Erhebungen richtig sind, sind wir berechtigt, eine Varianzanalyse nach dem Fischerschen Faktoriellen Schema durchzuführen. (Fig. 3)

Die Vorteile, die die Anwendung dieser Methode mit sich bringt haben wir schon früher kurz geschildert. Hier können wir uns damit begnügen, die Möglichkeit die Interaktion unter Kontrolle zu bringen, zu unterstreichen.

Ich habe es immer für sehr merkwürdig gehalten, dass die Gegner der Gestalttheorie die Ermittlung der Interaktion - in anderen Worten, die Möglichkeit, quantitativ zu beweisen, dass die gleichzeitige Wirkung mehrerer Bedingungen, verschieden als die Summe ihrer isolierten Bedingungen sei - für einen Beweisgrund gegen die Gestalttheorie hielten. Es scheint mir dagegen ein mächtiges Forschungsinstrument für die Gestaltforscher zu sein.

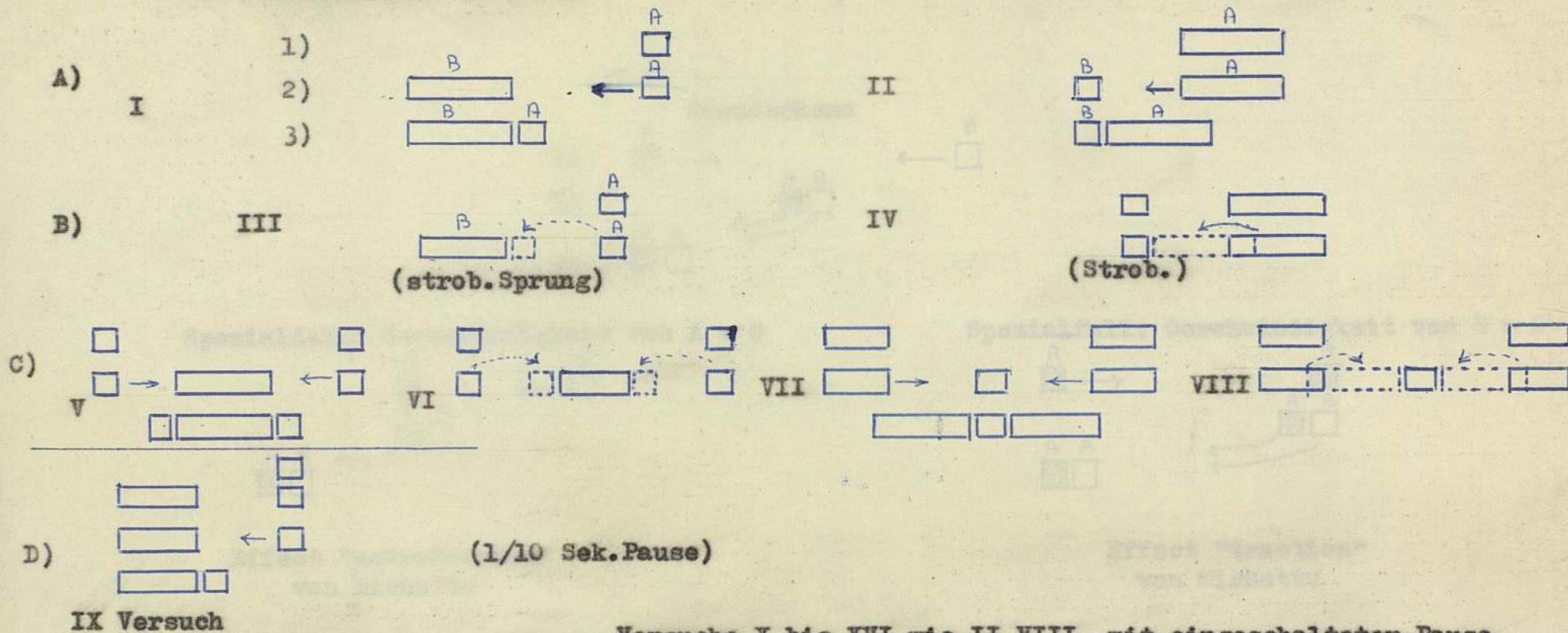
Aber auch abgesehen davon, scheint mir die Möglichkeit einer wohlgeplanten Experimentierung, die die Anwendbarkeit der varianzanalytischen Methodik mit sich bringt, die zweite Phase der Forschung im Wahrnehmungsgebiete wesentlich zu bereichern.

Noch möchte ich hinzufügen, dass, wo die planmässige Variierung der Bedingungen, die die Fischersche Methodik fordert, nicht durchführbar ist, die Faktorenanalytische Methodik Anwendung finden sollte.

Doch sollte man noch eine Stufe weitergehen. Bisher haben wir nur die Methodik geschildert, durch die die Wirkung der Bedingungen bewiesen sein soll. Die nächste Frage führt zu den Gesetzen dieser Wirkung, die uns einen tieferen Einblick in deren Natur verschaffen sollte. (Fig. 4). Es ist nicht schwer, nach den Fischerschen Modellen eine Versuchsserie zu planen, die uns den Gang der verschiedenen Veränderlichen zu ermitteln, und aus diesen die bezüglichen mathematischen Funktionen zu bestimmen erlaubt. Es gibt aber bedeutende

Schema

für die varianzanalyse der Bedingungen A) B) C) D) in einem qualitativen Wahrnehmungsexperiment (Phänomenale Anziehung)



IX Versuch

Versuche X bis XVI wie II-VIII, mit eingeschalteter Pause

Abhängige Veränderliche: Anziehungseindruck

1 = B zieht A an

0 = A bewegt sich unabhängig von B

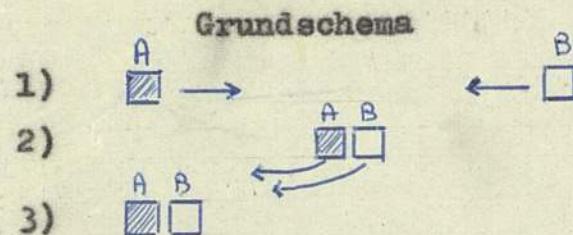
Unabhängige Veränderlichen  
(varierte Bedingungen)

- A) Grösse (+, -)
- B) Geschwindigkeit (+, -)
- D) Pause (1, 0)
- C) Zentralität (1, 0)

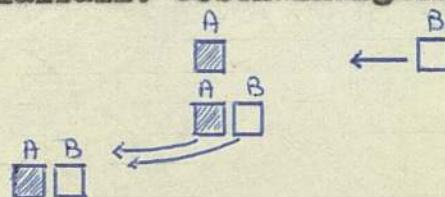
Zur Erläuterung: I Versuch: Zuerst ist an der Versuchsebene nur A zu sehen; dann erscheint plötzlich B, und A fängt gleich an sich gegen B langsam zu verschieben, bis er B erreicht.

Fig. 3

Funktionelle Analyse der Bedingungen in einem Wahrnehmungspsychologischen Experiment  
 (Phänomenales Schiebungs-Ziehungseffekt)

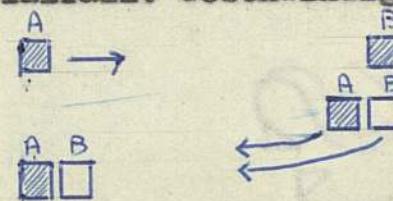


Spezialfall: Geschwindigkeit von A = 0



Effect "entrainment"  
 von Michotte

Spezialfall: Geschwindigkeit von B = 0



Effect "traction"  
 von Michotte

Abhängige Veränderliche: Ursacheseindruck

+= A zieht B

-= B schiebt A

0= A und B bewegen sich unabhängig

Unabhängige Veränderlichen:

- 1) Geschwindigkeit von A: wird in 5 Stufen variiert (0,1,2,3,4)
- 2) " " B: wird in 5 Stufen variiert (0,1,2,3,4)
- 3) " " AB: nach dem Kontakt: wird in 4 Stufen variiert (1,2,3,4)

Zahl der Experimente =  $5 \times 5 \times 4 = 100$

Versuchstechnische Schwierigkeiten, da sich zu diesem Zwecke die Zahl der Experimente vervielfachen muss.

Mit dieser Andeutung will ich meinen Vortrag enden. Nur möchte ich nicht den Eindruck hinterlassen, meine Meinung sei, dass eine mathematische Formel das Ende einer Untersuchung bedeute. Im Gegen teil, ist es nur wieder ein Anfang.

ERNST ZILKOVIC