

H. THOMÄ, H.-J. GRÜNZIG, H. BÖCKENFÖRDE UND H. KÄCHELE, ULM \*

## Das Konsensusproblem in der Psychoanalyse \*\*

*Übersicht:* Nach eingehender Diskussion methodologischer Fragen zur Konsensusproblematik in der Psychoanalyse (I. Teil), insbesondere unter Berücksichtigung der Argumentationen von Ricoeur und Lorenzer, und nach einer kritischen Literaturübersicht wird über zwei empirische Untersuchungen zu diesem Problemkreis berichtet. Einmal wurden 55 Stundenprotokolle einer psychoanalytischen Behandlung mit Hilfe von Schätzskalen nach operationalisierten Angstkonzepten (Beschämungsangst, Kastrationsangst, Schuldangst, Trennungsangst) eingestuft. Die drei Beurteiler erzielten hierbei mittlere Inter-Beurteiler-Übereinstimmungen, die mit den in der Literatur dokumentierten Werten in Einklang stehen. In der zweiten Untersuchung wurden die Gesamtstunden in einzelne Patientenäußerungen segmentiert; nach einem sorgfältigen Auswahlverfahren kristallisierten sich 40 dieser Äußerungen heraus, welche dann sowohl 25 Psychoanalytikern wie auch 45 Laien zur Beurteilung im Hinblick auf die Angstkonzepte vorgelegt wurden. Der Sachverhalt, daß Experten und Laien zu hohen Inter- wie auch zu hohen Intragruppen-Übereinstimmungen gelangten, weist darauf hin, daß die Operationalisierung dieser Konzepte im Beurteilungsprozeß ein theoriebezogenes Vorwissen entbehrlich macht. Insgesamt läßt sich feststellen, daß die Auswahl der Textsegmente von verschiedenen Gruppen bestätigt wurde, was als wesentliche Voraussetzung für eine erfolgreiche Konstruktvalidierung der Angstkonzepte angesehen werden kann.

### *I. Problemstellung und kritische Literaturübersicht*

Übereinstimmung und Widerspruch, Zustimmung und Ablehnung, Einverständnis und Meinungsverschiedenheit kennzeichnen das berufliche Leben des Psychoanalytikers in allen Bereichen. In der psychoanalytischen Behandlungssituation werden Interpretationen anhand ihrer therapeutischen Wirkung, am „Ja“ und „Nein“ des Patienten und — gleichsam in höchster Instanz — am Mitsprechen der Symptome, die das letzte Wort haben, erprobt. Hierbei ist die therapeutische Interak-

\* Aus der Abteilung für Psychotherapie (Leiter: Prof. Dr. H. Thomä) des Psychosozialen Zentrums der Universität Ulm.

\*\* Die vorliegende Untersuchung wurde mit Unterstützung der Deutschen Forschungsgemeinschaft durchgeführt. Sie wurde durch die Mitwirkung von Mitgliedern der Ulmer Psychoanalytischen Arbeitsgemeinschaft, der Abteilung für Psychotherapie und der Abteilung für Psychosomatik der Universität Ulm sowie von 25 Psychoanalytikern aus dem gesamten Bundesgebiet ermöglicht, denen wir für diese Unterstützung herzlich danken. An der Erhebung und Auswertung der Daten war Dipl.-Psych. Cornelia Schaumburg mitbeteiligt.

tion keineswegs eine private Angelegenheit. Indem der Psychoanalytiker eine theoriebezogene Methode vertritt, ist er dem wissenschaftlichen Kanon unterworfen: die im Sprechzimmer gemachten Beobachtungen und ihre Interpretationen werden in Fachzeitschriften publiziert. Schon auf der klinischen Diskussionsebene wird eine Öffentlichkeit unter Experten in diagnostischen oder behandlingstechnischen Seminaren hergestellt. Hier vollzieht sich eine kritische Auseinandersetzung, die sich z. B. um das Verständnis der Phänomene, um ihre Benennung und theoretische Einordnung dreht. Grundlegende Übereinstimmung über das methodische und theoretische Paradigma<sup>1</sup> Sigmund Freuds reicht als Entscheidungshilfe nicht aus. Es genügt nicht, allein im stillen Kämmerlein Widerstand und Übertragung zu beobachten, in ihnen unbewußte Motive zu erkennen und sie systematisch auf ihren Sinnzusammenhang mit Symptomen zu untersuchen. Die konsensuelle Validierung durch andere Psychoanalytiker stellt einen wesentlichen Schritt der Methoden- und Theorieprüfung dar. Nur durch Übereinstimmung unter Experten kann z. B. festgelegt werden, ob in einem Behandlungsbericht überhaupt Psychoanalyse (als Methode und Theorie) adäquat zur Anwendung kam (vgl. Schlesinger, 1974). Ebenso ist der Psychoanalytiker in seinem beruflichen Identitätsgefühl auf konsensuelle Validierung, auf stillschweigende oder ausgesprochene Bestätigung einer Gruppe angewiesen (vgl. dazu Thomä, 1976).

Gemeinsamkeiten hinsichtlich des grundlegenden psychoanalytischen Paradigmas treten in klinischen Diskussionen oft rasch in den Hintergrund. In behandlingstechnischen Seminaren wird in der Regel auch dort nicht vom Konsens ausgegangen, wo ein solcher vorliegt oder leicht erreicht werden könnte. Klinische Diskussionen scheinen eher die Funktion zu haben, die Übereinstimmung des Psychoanalytikers mit sich selbst und mit seinem Patienten, nämlich die Stimmigkeit seiner Deutungen in Frage zu stellen und alternative Auffassungen vorzuschlagen oder zu erwägen. Es handelt sich hierbei um ein zwangsläufiges Geschehen, denn bei einem erreichten Konsens, im Augenblick der völligen Übereinstimmung, fände die Diskussion ein Ende. Auf der anderen Seite brauchen verschiedene Deutungen dem erreichbaren Konsens nicht zuwiderzulaufen; Vieldeutigkeit und Überdeterminierung er-

<sup>1</sup> Unter „Paradigma“ versteht der Wissenschaftshistoriker Thomas Kuhn, „daß einige anerkannte Beispiele für konkrete wissenschaftliche Praxis — Beispiele, die Gesetz, Theorie, Anwendung und Hilfsmittel einschließen — Modelle abgeben, aus denen bestimmte festgefügte Traditionen wissenschaftlicher Forschung erwachsen“ (1967, S. 28).

lauben verschiedenartige Deutungen. Da Psychoanalytiker die Phänomene auf ihre Vieldeutigkeit hin betrachten, ergibt sich bereits auf der praktisch-klinischen Diskussionsebene das bekannte: „Ja, aber könnte es nicht auch so oder anders sein ...“

Hiermit ist ein wesentlicher Aspekt der Konsensusproblematik angesprochen, nämlich die Frage nach intersubjektiver Übereinstimmung auf der klinischen Diskussionsebene. Meinen Kliniker tatsächlich dasselbe, wenn sie denselben Begriff verwenden? Lassen sich klinische Konstrukte so scharf und prägnant formulieren und definieren, daß sie in identischer Bedeutung verwendet werden? Diese und viele andere sich hieran anschließende Fragen sind einer gründlichen empirischen Untersuchung keineswegs unzugänglich. Trotz der Berücksichtigung des individuellen Moments, der „persönlichen Gleichung“ des Klinikers, ist ein intersubjektiv anerkannter Gebrauch und ein allgemeinverbindlicher Inhalt der Begriffe mit S. Freud zu fordern, der im Hinblick auf die Bedeutung der Übereinstimmung folgendes ausführte: „Da man niemand seine Abnormität beweisen kann, wird eine allgemeine Übereinstimmung in den Dingen der Tiefenpsychologie besonders schwer zu erreichen sein. Manche Psychologen meinen sogar, dies sei ganz aussichtslos und jeder Narr habe das gleiche Recht, seine Narrheit für Weisheit auszugeben. Ich bekenne, ich bin hierin optimistischer. Unsere Erfahrungen zeigen doch, daß auch in der Psychologie ziemlich befriedigende Übereinstimmungen zu erreichen sind. Jedes Forschungsgebiet hat eben seine besondere Schwierigkeit, die zu eliminieren wir uns bemühen müssen“ (Freud, 1926 b, S. 250).

Im Hinblick auf die Konsensusproblematik ist dieser Empfehlung Freuds nur in seltenen Fällen Folge geleistet worden. Nicht wenige Kliniker fürchten eine Einengung des freien Spiels von Einfällen in der Falldiskussion, wenn sie mit Einstufungsskalen operieren und sich auf Beurteilungskonzepte festlegen lassen sollen. Diese Abneigung hängt mit der Funktion von kasuistischen Seminaren zusammen, die, wie wir bereits kurz erwähnten, das freie Spiel der Phantasie über mögliche Sinnzusammenhänge kultivieren müssen, um alternative neurosenpsychologische Interpretationen in Gedankenexperimenten zu erproben. Hierbei werden intuitive Vorgänge weder bezüglich ihrer Entstehung, noch im Hinblick auf ihre Begründung näher untersucht, und Entscheidungen über die Gültigkeit der einen oder anderen psychodynamischen Interpretation können erst in der Behandlungssituation getroffen werden. Sobald kasuistische Seminare eine größere Verantwortung in Diagnostik oder Indikationsstellung tragen, wird es unerlässlich, Entscheidun-

gen zu begründen und eine Übereinstimmung unter Experten herbeizuführen. Als Beispiel sind Seminare zu nennen, die der Bestimmung eines psychodynamischen Behandlungsfokus dienen. Dort wird gefordert, die psychodynamischen Thesen anhand der beobachtbaren Phänomene, besonders anhand der Übertragungs- und Gegenübertragungsformen, zu begründen und eine Klärung auf der Theorieebene herbeizuführen. Dieses Festlegen beendet zunächst das freie Spiel von Einfällen, macht sie verbindlich und bringt sie der Stunde ihrer Wahrheit näher.

Es kann keinen Zweifel daran geben, daß die intersubjektive Übereinkunft unter Experten, die konsensuelle Validierung eine der Voraussetzungen wissenschaftlicher Objektivität ist. Wie groß die Meinungsverschiedenheiten unter Vertretern verschiedener Wissenschaftslehren auch sonst sein mögen, so scheint von keiner Seite bestritten zu werden, daß zumindest der Kampf um die intersubjektive Verständigung bis zur bestmöglichen Konsensbildung in der Experimentier- und Interpretationsgemeinschaft der Wissenschaftler einen hervorragenden Platz einnimmt (Apel, 1976 a, 1976 b; Popper, 1958; Taylor, 1975). Es wird anerkannt, daß Untersuchungen über Entstehung und Begründung von Konsens und Dissens als Korrektiv eines solipsistischen Wahrheitsverständnisses unerlässlich sind. So läßt sich die oben zitierte Bemerkung von Freud auch als Forderung nach empirischen Untersuchungen auffassen, durch die die „Narrheit“ eines solipsistischen Wahrheitsanspruches korrigiert werden kann. Die bisherigen Ausführungen zeigen, daß sich Übereinstimmungen auf verschiedene Bereiche beziehen können, auf theoretische Annahmen und beobachtbare Phänomene. Die Bezeichnung „Inter-Beobachter-Übereinstimmung“ könnte nahelegen, daß sich Übereinstimmungsuntersuchungen nur auf beobachtbare vorhandene Phänomene beziehen. Nun wird im Anschluß an Ricoeur besonders von Lorenzer betont, daß die Psychoanalyse weder eine Verhaltens- noch Beobachtungswissenschaft sei. Ricoeur erweckt den Anschein, als würde das Konsensusproblem in einer rein hermeneutischen Psychoanalyse irrelevant. Daß schließlich aber auch Ricoeur und Lorenzer nicht umhin können, die Übereinstimmung unter Experten als wesentliches Kriterium für die Richtigkeit praktischen und wissenschaftlichen Tuns zu implizieren, wollen wir im folgenden anhand einiger repräsentativer Textstellen demonstrieren.

Betrachten wir zunächst einige Stellen aus dem letzten Werk Ricoeurs, in dem er ausführlich auch Stellung zur Kritik an seinem früheren Freud-Buch „Die Interpretation“ (1965) nimmt. In „Hermeneutik und Psychoanalyse“ (1969) heißt es,

„es gehe der Psychoanalyse weder um ‚Tatsachen‘ noch um die Beobachtung von ‚Tatsachen‘, sondern um die Interpretation einer ‚Geschichte‘; selbst die von außen beobachteten und im Laufe der Analyse berichteten Tatsachen haben nicht als Tatsachen ihre Bedeutung, sondern als Äußerungen der Sinnwandlungen, die sich in dieser Geschichte ereignen. Die Verhaltensänderungen werden nicht als ‚Observablen‘ gewertet, sondern als ‚Signifikanten‘ der Geschichte des Wunsches. Demnach hat die Psychoanalyse ihren eigenen Gegenstand stets in Sinneffekten — Symptomen, Wahngebilden, Träumen, Illusionen —, die die empirische Psychologie nur als Verhaltenssegmente betrachten kann; der Analytiker jedoch behandelt gerade das Verhalten als ein Segment des Sinnes“ (S. 114).

Man müsse deshalb den Attacken der Logiker, Semantiker und Wissenschaftstheoretiker, welche der Psychoanalyse den Charakter einer Wissenschaft abstreiten, alles zugestehen und dieses Zugeständnis in eine Entgegnung verwandeln. Die Differenz zum Behaviorismus bestehe von Anfang an und sei eine totale. Die Psychoanalyse setze (im Gegensatz zum Behaviorismus) nicht bei beobachtbaren Verhaltensweisen an, sondern beim Unsinnigen, das nach Interpretation verlange. Demgemäß verfehle jeder Versuch, die Psychoanalyse der Beobachtungswissenschaft oder einer ihr entspringenden Technik anzugleichen, das Wesentliche.

Ricoeur scheint festlegen zu wollen, daß sich „Beobachtungs- und Verhaltenswissenschaften“ auf interpretationsfreie nackte Tatsachen der unbelebten Natur, auf „data bruta“ (Taylor) richten. Dort bringt Ricoeur die empirisch-wissenschaftliche Psychologie unter, die er mit dem Behaviorismus gleichsetzt. Die Psychoanalyse wird bei den interpretierenden Sozial- und Humanwissenschaften eingeordnet. Durch diese extreme Aufspaltung wird Zusammengehöriges so zerrissen, daß nach Auffassung Ricoeurs eine „sinnentleerte“ wissenschaftliche Psychologie und eine hermeneutische Psychoanalyse entstehen. Die eine Psychologie befaßt sich nach Ricoeur mit beobachtbaren, aber sinnleeren Verhaltensweisen, die andere sucht im „Unsinnigen“ von Symptomen, Wahngebilden, Träumen und Illusionen den (unbewußten) Wunsch. Wie dieses „Unsinnige“ aber gegeben ist, läßt Ricoeur offen (1969, S. 113). Es dürfte kaum anders als in der Selbst- oder Fremdbeobachtung unsinniger Verhaltensweisen auftreten. Symptome, also unsinnige Verhaltensweisen sind ebenso beobachtbar wie introspektiv gewonnene Mitteilungen eines Patienten über sich selbst, seine ihn quälenden Gedanken und seine „freien“ Assoziationen dazu. Freud hat in solchen unsinnigen Erscheinungen einen Sinn entdeckt und darunter „Bedeutung, Absicht, Tendenz und Stellung in einer Reihe psychischer Zusammenhänge“ (1916/17, S. 55) verstanden. Diese methodenunabhängige Beobachtbarkeit psychopathologischer Phänomene ist klinisch und wissenschaftlich

hervorzuheben. Unsinnige Symptome werden vorgefunden und keineswegs hergestellt. Ricoeur behauptet demgegenüber, daß psychoanalytische und physikalische „Gegenstände“ jeweils durch die hermeneutische bzw. naturwissenschaftliche Methode erst konstituiert würden.

Beide „Gegenstände“, das Unbewußte der Psychoanalyse und der physikalische Gegenstand, existieren nach Ricoeur nur im relativen Verhältnis zur Methode:

„Wir müssen das Unbewußte relativieren, andererseits aber unterscheidet sich diese Relativität nicht von jener des physikalischen Gegenstandes, der seine ganze Realität aus der relativen Beziehung zur Gesamtheit der wissenschaftlichen Schritte, die ihn konstituieren, empfängt. Die Psychoanalyse hat Anteil am selben ‚angenäherten Rationalismus‘ wie die Naturwissenschaft“ (1969, S. 18).

Schließlich endet auch Ricoeur in seiner Beweisführung bei der Beobachtung, indem er sagt: „Die Gewähr dafür, daß die Realität des Unbewußten nicht nur ein reines Hirngespinnst der Psychoanalyse ist, *erbringt uns schließlich allein der therapeutische Erfolg*“ (1969, S. 19, von uns hervorgehoben). Wie sollte aber der Kliniker den therapeutischen Erfolg feststellen können, wenn unsinnige Verhaltensweisen weder vom Patienten selbst noch von Dritten beobachtet werden könnten? Daß die beobachtbaren Verhaltensänderungen, Besserungen oder Heilungen eines Patienten in ihrer Abhängigkeit von der interpretativen Technik ihrerseits wieder interpretiert werden können und müssen, ändert nichts an der Tatsache, daß sie als solche vom Patienten selbst und von Dritten beschrieben werden können. Schließlich ist bei Nachuntersuchungen eine Befunderhebung durch mehrere, voneinander unabhängige Beobachter und deren Übereinstimmung bzw. Nichtübereinstimmung entscheidend (R. Argelander et al., 1976).

H. Göttner (1973) hat von Verwirrungen gesprochen, die der vieldeutige Ausdruck, daß eine Hypothese ihren Gegenstand erst konstituiert, hervorruft:

„Schließlich besagt er nicht mehr, als daß die Hypothese ihre Endhypothesen bestimmt und diese wiederum ihren Objektbereich. Aber diese Bestimmung der Relevanz der einzelnen Felder des Objektbereichs für eine Hypothese besagt ja nicht, daß diese Erwartungen durch empirische Befunde in diesem Felde erfüllt werden, sie sagt lediglich etwas über die Rangfolge aus, welche die anschließend gemachten empirischen Befunde haben werden.

Das methodologische Verfahren der Naturwissenschaften unterscheidet sich hiervon nicht: Auch der Naturwissenschaftler beobachtet seine Objekte nicht an sich und bildet danach Hypothesen, sondern auch für ihn ist die Theorie das erste, und Beobachtungen macht er in Experimentsituationen, deren Richtung von dieser Theorie bis ins Detail veranlaßt ist. Ohne diese Theorie wäre er genau wie der Literaturwissenschaftler außerstande, etwa gemachte Beobachtungen *überhaupt zu deuten*, erst die

Theorie *bringt die Beobachtung ‚zum Reden‘*. Das Entscheidende aber ist, daß auch hier die Beobachtungen zwar unter Anleitung der Theorie gemacht werden, in ihrer Überprüfung aber unabhängig von dieser sind, denn sonst wäre die Widerlegung von Hypothesen unmöglich“ (S. 36).

Ricoeurs dogmatische Aufteilung in eine wissenschaftliche, behavioristische Psychologie, die beobachtet, aber nichts Sinnvolles sieht, und in eine hermeneutische Psychoanalyse, die den Sinn im Unsinn deutet, ohne ihn tatsächlich zu beobachten, wurde grundlegend von H. Kunz kritisiert, der zu folgendem Ergebnis kommt:

„Ricoeurs Aneignung der Psychoanalyse gleicht dem Unterfangen eines Menschen, der die Pflanzen eines Landes allein aufgrund des Studiums einer entsprechenden Flora, d. h. der in ihr enthaltenen Beschreibungen und Abbildungen, kennenlernen will, sich jedoch nicht um die in der Natur vorkommenden pflanzlichen Individuen kümmert, auf die sich letztlich die Diagnosen, Abbildungen, Verbreitungsangaben usw. beziehen. Er kann dann die Namen der Arten und ihre systematische Ordnung beliebig ändern und schließlich die ‚Buch-Flora‘ mit der ihm unbekannten wirklichen naturwüchsigen Flora identifizieren“ (H. Kunz, 1975, S. 50/51).

Zu einem ähnlichen Ergebnis wie Kunz gelangen Ramzy und Shevrin (1976), deren Arbeit sich mit dem Problem des Schlußfolgerns (inference) bei psychoanalytischen Interpretationen befaßt. Ohne Ricoeurs Beitrag zur Rolle der Sprache und der Symbolik in der Psychoanalyse zu schmälern, kritisieren diese Autoren seinen Nominalismus und betonen, daß Ricoeur die wissenschaftlichen Probleme hinsichtlich der Interpretation noch weiter durcheinander gebracht hat. Ihr Einwand lautet:

„Es trifft nicht zu, daß der Psychoanalytiker nur von der Sprache, so wesentlich sie auch ist, abhängt. Er ist auch ein Beobachter des Gesamtverhaltens des Patienten. Er befindet sich nicht einfach in der Lage eines Gelehrten, der nicht mehr in Händen hat als einen Text, sondern der Psychoanalytiker ist gegenwärtig, wenn ein ‚Text‘ entsteht, eine Situation, die dem Bibelexegeten nicht vergönnt ist. Zum anderen ist die Kommunikation als solche Gegenstand der Beobachtung. Die Äußerung eines Menschen über seine Erfahrung ist nicht nur ein zu interpretierender Text, sondern ein Faktum für sich selbst. Sie ist ebenso eine Tatsache wie ein Stern oder eine Blume<sup>2</sup>. Doch noch gewichtiger ist ein weiterer Einwand: wenn nämlich der Psychoanalytiker mit der Sprache von Wünschen zu tun hat, dann muß es Wege geben, wie diese Wünsche in Sprache verwandelt werden. (...) Ist das denn nicht ganz genau das besondere psychoanalytische Problem: Wie kommt ein Mensch mit seinen Sehnsüchten (Wünschen, Trieben, Impulsen usw.) zurecht, die er als wirklich und keineswegs nur als Gedanken über Ereignisse erlebt?“ (S. 157).

Diese Argumente gegen eine reine Hermeneutik in der Psychoanalyse haben wir in ähnlicher Weise an anderer Stelle vorgebracht (Thomä und Kächele, 1973).

<sup>2</sup> Vgl. dazu das Flora-Gleichnis von H. Kunz.

Nach Freuds epistemologischem Selbstverständnis ist die Psychoanalyse eine Wissenschaft, die in der „Deutung der Empirie“ gründet (S. Freud, 1914, S. 142). Auch wenn Freuds Aussagen zur Wissenschaftslehre und Psychoanalyse unterschiedliche Auslegungen zulassen, so scheint nach den Zeugnissen der Fachliteratur Übereinstimmung hinsichtlich dieser Grundlegung — „Deutung der Empirie“ — zu bestehen. Umstritten sind sowohl allgemeine neurosentheoretische Interpretationen wie die auf diese bezogenen technischen Deutungen. Um Meinungsverschiedenheiten klären zu können, ist es wesentlich, ob Beobachtung und Deutung so ineinander übergehen, daß weder in der Selbst- noch in der Fremdbeobachtung eine Unterscheidung zwischen Beschreibung und Bedeutungsverleihung getroffen werden kann. Schon die Fähigkeit, in Distanz zu den beobachteten Phänomenen treten und ihnen eine unterschiedliche Bedeutung verleihen zu können, zeigt an, daß im Erleben Wahrnehmung und Interpretation gedanklich getrennt werden können. Tatsächlich liegen die Konsensusprobleme nicht in der Deskription der Phänomene, sondern in ihrer psychodynamischen Zuordnung unter Gesichtspunkten der Ähnlichkeit, also mit Blick auf Verallgemeinerungen und Regelmäßigkeit.

Die Praxis fordert im Interesse der Patienten Konsensusuntersuchungen. Dieser Notwendigkeit kann sich auch Ricoeur nicht entziehen, und sie hebt seine dogmatische Aufteilung auf, auch wenn er diese Konsequenzen nicht gezogen hat: Wenn allein der therapeutische Erfolg die Gewähr dafür bietet, daß die Realität des Unbewußten nicht nur ein reines Hirngespinnst ist, wie Ricoeur feststellte (1969, S. 19), dann sind entsprechende Untersuchungen vordringlich. Der von Ricoeur geforderte Gültigkeitsnachweis setzt aber Konsensusbestimmungen voraus: „Als Prüfstein läßt sich durchaus die von der Erkenntnistheorie gestellte Forderung verwenden, ein *standardisiertes Material klinischer Tatsachen* von einer bestimmten Anzahl voneinander unabhängiger Forscher auswerten zu lassen“ (Ricoeur, 1965, S. 383; von uns hervorgehoben). Unter welchem wissenschaftstheoretischen Vorzeichen man die Dinge auch betrachten mag, uns ist hier allein wesentlich, daß schließlich auch Ricoeur den Beweis für unbewußte Motive in therapeutischen Veränderungen sucht, die als Tatsachen konsensuell feststellbar sind. Gerade weil es sich um Sinnbezüge handelt, mit denen der Psychoanalytiker zu tun hat, und weil keine allgemeingültigen Interpretationsregeln vorliegen (Rapaport, 1960; Eissler, 1958; Ramzy, 1974), schaffen Untersuchungen, die der Entstehung von Übereinstimmungen und ihrer Begründung dienen, Abgrenzungskriterien zum Un-



sinn und zur „Narrheit“. Deshalb kann schließlich auch Lorenzer (1974) nicht umhin, der Konsensusbildung in seinem System einen hervorragenden Platz einzuräumen. Anscheinend in Anspielung auf Glover<sup>3</sup> heißt es:

„Die Achillesferse der Psychoanalyse als Wissenschaft ist die ‚kasuistische Diskussion‘... Das Verhältnis zur Theorie ist kompliziert. Die *Konsensbildung*<sup>4</sup> verläuft zunächst unterhalb der Ebene der Theoriediskussion im Versuch, die einzeln analytisch profilierten Vorannahmen ins horizontal und vertikal ausgespannte Netz der typischen lebenspraktischen Vorannahmen konsistent einzufügen. Das einzeln analytisch bestätigte Interaktionsmuster und die typischen Vorannahmen müssen jene strukturelle Übereinstimmung zeigen, die das System aller Vorannahmen auszeichnet. Die Merkmale ihrer konkreten Ausgestaltung dürfen den typischen Vorannahmen des ersten Abstraktheitsgrades nicht so widersprechen, daß der Bedeutungsrahmen gesprengt wird“ (S. 168, Hervorhebung von uns).

Und an anderer Stelle:

„Wenn der Analytiker das Erscheinungsbild einer bestimmten Interaktionsstruktur als ‚hysterisch‘ im Seminar darstellt, so reklamiert er nicht in erster Linie die Kenntnisse der anderen Seminarmitglieder über theoretische Zuordnungen, sondern er beansprucht die Vorstellung von typischem hysterischem Interagieren. Die Vorannahmen stellen das unabdingbare Vermittlungsglied der Diskussion dar. Nicht von ungefähr läuft die Konsensusbildung auch hier nicht über die mehr oder minder geglückte Einpassung in den Katalog theoretischer Kriterien, sondern immer über die Frage, inwiefern das präsentierte Bild einer analytischen Episode sich den Vorannahmen der anderen in einer ihrer Erfahrung entsprechenden Weise fügt. Wenn die Szene z. B. von den einen Seminarteilnehmern als ‚ödipal‘, von den anderen als ‚präödipal‘ interpretiert wird, so geht es — anders als bei internistischen Falldiskussionen — nicht um den Vergleich von Daten und Hypothesen, sondern um die *Stimmigkeit der Gestalt*, was heißt: Es geht um die Vermittlung der geschilderten Szene an die reklamierten szenischen Vorannahmen der Beteiligten“ (S. 163/164).

Würde man daran gehen, aus dieser Darstellung eine empirisch-hermeneutische Untersuchung aufzubauen, würde man u. E. unweigerlich mit

<sup>3</sup> „Die Achillesferse der psychoanalytischen Forschung“ besteht nach Glover (1952) darin, „daß in jedem Fall die Interpretation ein wesentlicher Teil des psychoanalytischen Untersuchungsprozesses ist, es aber keine wirksame Kontrolle über die Schlußfolgerungen gibt, die aus Interpretationen abgeleitet werden“ (S. 405). Dieses Zitat läßt erkennen, daß Lorenzer eine andere Achillesferse der Psychoanalyse beschreibt als Glover. In der kasuistischen Diskussion die Wissenschaftlichkeit der Psychoanalyse zu verankern bzw. dort zu lösen, wirft noch größere Probleme auf, auf die wir noch eingehen werden und die empirische Konsensusforschungen um so dringlicher machen.

<sup>4</sup> Der Konsens unter Experten ist ein wesentlicher Schritt auf dem Weg zur wissenschaftlichen Validierung, aber noch nicht diese selbst. Lorenzer verkleinert allerdings zugleich die Bedeutung des Konsenses, wenn er ihn auf eine ideologieresistente Forscher- und Interpretationsgemeinschaft beschränkt, denn er unterstellt hierbei, daß Experten keine Bereitschaft haben, sich selbst in Frage zu stellen, obwohl sicherlich eine natürliche Neigung besteht, im Rahmen der jeweiligen wissenschaftlichen Annahmen Untersuchungen durchzuführen. Im übrigen ist gerade deshalb eine Objektivität zu fordern, die auch unabhängigen Wissenschaftlern eine Überprüfung und Urteilsbildung und Einsicht in Begründungszusammenhänge erlaubt.

Problemen konfrontiert werden, mit denen alle Untersucher jenseits der wissenschaftstheoretischen Lager zu ringen haben. Wissenschaftstheoretische Bekenntnisse helfen nicht weiter, wenn Aussagen über „Stimmigkeiten von Gestalten“, über „strukturelle Übereinstimmung“, über „typische Interaktionsstrukturen“ gemacht und auf ihren Konsensus hin verglichen werden sollen. Lorenzer kann den entsprechenden Problemen entgehen, da er empirische Untersuchungen über seine zentralen Thesen bisher nicht vorgelegt hat. Er würde gewiß kaum daran festhalten können, daß „die Konkurrenz der Forscher (...) im hermeneutischen Feld der Analytikergruppe“ (S. 194) bei solchen Untersuchungen beseitigt würde.

Tatsächlich scheint die „Bornierung“ bisheriger psychoanalytischer Konsensusuntersuchungen in Lorenzers Sicht darin zu liegen, daß sie nicht von seinem in sich ziemlich geschlossenen System ausgehen. In ihm wird in einem vielgliedrigen Prozeß durch „Sprachspiele“ die Funktion der Wirklichkeitsprüfung wahrgenommen. Eine wesentliche Rolle bei der Übereinstimmung von Wirklichkeit und Theorie spielt die „interaktionistisch perspektivierte Metapsychologie“ (S. 159), die auch als „Sprachbaum“ vorgestellt wird. Lorenzer betont den „qualitativen Sprung“ (S. 170), der zur abstraktesten Stufe allgemein-typischer Sprachspiele — Lorenzers Metapsychologie — führe, und stellt fest, daß jede Theorie nur insofern Wert habe, als vom abstrakten Modell unablässig zum „konkreten“ aufgestiegen werde (S. 89). Gerade wegen des hohen Abstraktionsniveaus hat S. Freud die Metapsychologie als „spekulativen Überbau“ bezeichnet<sup>5</sup>. Wie man nämlich von der Abstraktion zum Konkreten, von der Theorie- zur Praxisprache gelangt, und wie überhaupt die metapsychologischen Gesichtspunkte in der psychoanalytischen Praxis vorkommen, ist ein offenes empirisches und epistemologisches Problem. Ebenso wenig ist geklärt, welche Teile der Metapsychologie indirekt empirisch in der psychoanalytischen Situation untersucht werden können. So wird die Metapsychologie bei Lorenzer

<sup>5</sup> Wie immer man Freuds Auffassung der Metapsychologie auslegen mag, so ist davon auszugehen, daß Freud und nicht Thomä und Kächele (1973) die Metapsychologie als „spekulativen Überbau“ bezeichnete. Lorenzer hat uns nämlich falsch zitiert, obwohl wir (S. 230) die Formulierung „spekulativer Überbau“ als von Freud stammend gekennzeichnet hatten. Wir haben auch keine globale Abwertung der Metapsychologie vertreten. Allerdings spricht aus unserer früheren Veröffentlichung eine Unsicherheit über die verschiedenen Dimensionen der Metapsychologie und ihre Bedeutung für die klinische Theorie der Psychoanalyse. Diese Unsicherheiten zu klären, steht gegenwärtig im Mittelpunkt der Theoriediskussion in der Psychoanalyse (vgl. z. B. G. S. Klein, 1976; Schafer, 1973; Friedman, 1976).

einerseits zum Meßresultat der Interaktionsformen (S. 191), wobei angeblich die metapsychologisch begriffenen typischen Vorannahmen jenen Vorannahmen, die im hermeneutischen Feld der Analyse eingesetzt werden, korrespondieren (S. 189). Andererseits könne die Metapsychologie auch als Meßskala angelegt werden (S. 172), und innerhalb ihres Rahmens können dann die Meßresultate noch zusätzlich vermessen werden (S. 189, S. 177): „Die Sprachspiele werden mittels des physikalistischen Begriffsrepertoires ‚vermessen‘. (...) Dieses ‚Vermessen‘ erfolgt auf allen Stufen der Abstraktion.“ Schließlich wird das Reden von Vermessen, Meßskala etc. zur Metapher (S. 195). Mit Hilfe der metaphorischen Verwendung des metapsychologischen Vermessens wird nun auf einmal das Energieverteilungsmodell als geradezu vorzüglich tauglich bezeichnet (S. 191), obwohl gerade daran das „szientistische Selbstmißverständnis“ der Psychoanalyse von Habermas und von Lorenzer demonstriert worden war. Hierbei würde sich nach unseren Erfahrungen erweisen, daß das metaphorische Vermessen anhand der von Lorenzer veränderten Metapsychologie erneut die zentrale Frage nach der „Wahrheit der psychoanalytischen Erkenntnis“ aufwirft.

Im Zentrum von Lorenzers Argumentation stehen Übereinstimmungen auf verschiedenen Ebenen. So heißt es beispielsweise: „Die strukturelle Übereinstimmung von bewährten (lebenspraktischen, Ref.) Vorannahmen und dem (theoretischen, Ref.) Modell wird als Erlebnis der stimmigen Gestalt bestätigt“ (S. 184). Obwohl nach Lorenzer die Achillesferse der Psychoanalyse als Wissenschaft in der kasuistischen Diskussion liegt (S. 167), wurden diesbezügliche Untersuchungen über strukturelle Übereinstimmungen oder über Stimmigkeit und Unstimmigkeit von Gestalten und die diese bestätigenden Erlebnisse von Lorenzer nicht durchgeführt. Da u. E. bei diesen „Gestaltwahrnehmungen“ jeweils Annahmen über noch nicht gestalthafte, unbewußte Motive beteiligt sind, unterliegen jene Vorgänge der Gefahr der „Hypergnosie“, wie Waelder (1926) die „Neigung, mehr Gestalt in der Realität zu sehen“, bezeichnet hat (1970, S. 290). Es geht hierbei um folgendes Problem: Wo Meßdaten an einer Skala abgelesen werden können, und wo es sich um die Wahrnehmung einfacher Sinnesdaten handelt, ist praktisch völlige Übereinstimmung zu erwarten. Doch schon einfache kombinatorische oder interpretative Urteilsleistungen, z. B. in der Röntgendiagnostik des Magengeschwürs (vgl. Hoffman et al., 1968; Goldberg, 1968), bringen erhebliche Probleme mit sich. Sehr viel komplizierter noch sind die Verhältnisse bei der Wahrnehmung von komplexen psychischen Phänomenen als Gestalten oder Sinnbezüge. Waelder (1970) bemerkt

hierzu: „Es werden mehr Menschen darin übereinstimmen, daß an irgendeiner Stelle etwas Rotes sichtbar ist, als in der Beobachtung von Haß oder Angst“ (S. 289). In Psychologie und Psychoanalyse gilt es einen Weg zu finden zwischen der Scylla des „Zuviel“ — der Hypergnosie in der Wahrnehmung bzw. Herstellung von Gestalten und Sinnbezügen — und der Charybdis des „Zuwenig“ — der Reduzierung von Wahrnehmungsprozessen auf exakt erhebbare, aber womöglich bedeutungslose Meßdaten. In diesem Zusammenhang führt Waelder (1970) Karl Mannheim an, der bereits 1929 den Sozialwissenschaftlern vorwarf, daß diese dazu neigen, das wichtig zu nehmen, was eben meßbar ist, und zwar nur deshalb, weil man es messen kann, anstatt das mit dem größtmöglichen Maß an Genauigkeit zu entdecken, was die größte Bedeutung hat.

Waelder läßt keinen Zweifel daran, daß in der Psychoanalyse eine solche größtmögliche Präzision angestrebt und auch erreicht werden kann, und zwar deshalb, weil Freuds Methode Einsicht mit objektiver Bestätigung verbindet. Der Psychoanalytiker prüfe seine Wahrnehmungen mit Hilfe von Beobachtungsdaten<sup>6</sup>, die nicht erst aus Schlußfolgerungen hervorgehen, sondern z. B. als Symptome vorweg da sind, wenn sie auch unterschiedlich benannt werden können. Deshalb können bereits auf dieser Ebene der Datenbenennung, die durch Konvention geregelt wird, Konsensusprobleme auftauchen. Der Psychoanalytiker oszilliere zwischen Gestaltwahrnehmungen und ihrer Prüfung mit Hilfe von Nichtgestaltoperationen. Die empirische Basis hierfür liegt in der langen Beobachtung von Patienten. Bei diesen Beobachtungen werden regelmäßige Verhaltensmuster und deren bedingte Veränderungen festgestellt, was induktiv Verallgemeinerungen und die Formulierung empirisch-statistischer Gesetze erlaubt. Solche Verhaltensmuster nennt Waelder auch Gestalten (1970, S. 315, 319). Obwohl der teilnehmende Beobachter bei der Registrierung solcher Verhaltensmuster zumindest in seiner Einsicht und Interpretationsfähigkeit berührt wird, kann hier von Beobachtungsdaten gesprochen werden, denn zumindest können unterschiedliche Interpretationen in der Schwebe gehalten werden, auch wenn es sich bei diesen Beobachtungsdaten nicht um „ganz nackte Tatsachen“ im Sinne von Taylors „data bruta“ handelt, „die ohne jegliche persönliche Einsicht oder Interpretation auf seiten des Beobachters verfügbar sind“ (1975, S. 260). Auch Taylor kennt data bruta verschiede-

<sup>6</sup> Vgl. hierzu S. Freud (1916/17, S. 276): „Ich habe diese Dinge (Symptome, Fehlleistungen etc., Ref.) nicht gemacht, sondern bloß gedeutet.“

ner Nacktheitsgrade, und er betont in diesem Zusammenhang die Bedeutung der intersubjektiven Übereinkunft für wissenschaftliche Objektivität. Demgemäß können sich Konsensusuntersuchungen auf verschiedene Ebenen der Betrachtung beziehen. Sie werden allerdings dann gegenstandslos, wenn die Übereinstimmung eines Autors mit sich selbst so gering ist, daß erst einmal die Widersprüche durch immanente Theoriekritik überwunden werden müßten, um sinnvolle Fragestellungen zu ermöglichen. Abgesehen vom immanenten Widerspruch sind einige Thesen Lorenzers so angelegt, daß sie jeder empirisch-psychoanalytischen Untersuchung und insbesondere den von Lorenzer zu Recht geforderten Konsensusuntersuchungen bei kasuistischen Diskussionen den Boden entziehen. Wir können im folgenden nur auf einige Ungereimtheiten dieser Art und ihre Folgen hinweisen:

S. Freud erklärte Verhaltensstörungen durch seelische Ursachen und nannte die Erklärungshypothesen Deutungen, die durch Ereignisse und Änderungen inner- und außerhalb der psychoanalytischen Situation bestätigt oder widerlegt werden können (S. Freud 1916/17, S. 45 und 52). Lorenzer führt einen wortreichen Kampf gegen „bornierte“ Psychoanalytiker — sein bevorzugtes Epitheton ornans für Kollegen —, die, von Verhalten und objektiven Symptomen ausgehend, in naiver oder auch epistemologisch durchdachter Weise nach Motiven als Gründen oder Ursachen jener Beobachtungsdaten suchen und ihre Hypothesen an wiederum intersubjektiv beobachtbaren Änderungen prüfen. Denn seine Thesen richten sich ebenso gegen Beobachtung wie gegen die Registrierung von Verhalten und auch gegen motivationstheoretische Erklärungen. Ursachen ebenso wie Objektivität entziehen sich u. E. nur dann psychoanalytischen Untersuchungen, wenn sie durch Lorenzers willkürliche Definition in die Gesellschaft bzw. in die historisch-materialistische Theorie verlagert werden: als äußere Ursachen und Ereignisse. Deshalb wird die Strukturanalyse einschneidend umbestimmt und der genetische Gesichtspunkt wird seines kausalanalytischen Anspruchs entkleidet (Lorenzer, 1974, S. 195).

Wie unhaltbar diese Aufteilung ist, zeigt sich daran, daß ihr nicht einmal Lorenzer selbst folgen kann. In seinen Sprachspielen kommt er ohne Motiv, objektives Symptom, Verhalten etc. nicht aus. Wie bei der Wiederkehr des Verdrängten schleichen sich Redensarten in Lorenzers System ein, die eigentlich exkommuniziert wurden. Wir geben einige Sprachproben: „Angriffspunkt der psychoanalytischen Untersuchung ist allemal eine in zwischenmenschlicher Interaktion und Kommunikation auftauchende Unerträglichkeit, sind Widersprüche wie der zwi-

schen Handlung und Motivation (was zumeist dem Patienten selbst verborgen bleibt), zwischen Wunsch und Erfüllung (was zunächst vom Patienten hingenommen werden muß) oder zwischen intendierter Darstellung und dem tatsächlichen Ausdruck — was objektiv als Symptom, in subjektiver Erfahrung aber als Leiden vom Patienten wie vom Untersucher ausgemacht werden kann“ (S. 214).

Als Ziel des psychoanalytischen Verfahrens beschreibt Lorenzer in guter Tradition die praktische Änderung, um an anderer Stelle hervorzuheben, daß „die praktisch ändernde Wirkung nicht den Rang einer beobachtbaren Datenerhebung habe“ (S.192): die Änderung könne nicht objektiviert werden. Im Widerspruch hierzu heißt es an anderer Stelle, daß Klischees (kurz: Verdrängtes) als Relikte das Verhalten bestimmen (S.198). In diesem alten psychoanalytischen Postulat ist die Voraussage enthalten, daß die Aufhebung der Verdrängung an verändertem Verhalten objektiviert werden kann. Gegen seine eigenen Thesen hat sich Lorenzer auch schon früher ausgesprochen, indem er von verhaltensmotivierenden unbewußten Inhalten sprach (1970, S. 134).

Schließlich kann man unseres Erachtens auch nicht ohne die Registrierung von Fakten auskommen, denn Lorenzers kritische Hermeneutik der „Destruktion des Faktisch-Gegebenen“ enthält Urteilsakte, die auf die Beobachtung nicht-destruierter Fakten angewiesen sind (1974, S. 277).

Gerade weil die psychoanalytischen Theorien letztlich ihre Proben in der psychoanalytischen Dyade bestehen und wegen der Rolle der Deutungstechnik hierbei, ist in kasuistischen Diskussionen wesentlich, daß die Verhaltensbeschreibung nicht übersprungen wird. Lorenzer schreibt, daß in der Psychoanalyse (als kritisch-hermeneutischer Wissenschaft) Daten nicht im Licht einer Theorie gesammelt, sondern deutend aufgelöst würden (S.277). Zweifellos sammeln wir Daten nicht als Selbstzweck. Schon die Art und Weise ihres Aufgreifens und Registrierens kann einen verändernden Charakter haben und neue Gesichtspunkte einführen. Therapeutisch gesehen wollen wir möglichst kurz beim Beschreiben und Klassifizieren der seelischen Erscheinungen stehen bleiben, aber die „wahrgenommenen Phänomene“ können in einer *dynamischen Auffassung* „gegen die nur angenommenen Strebungen“ erst „zurücktreten“ (S. Freud, 1916/17, S.62), wenn sie — im Lichte von Theorien, aber nicht durch sie hergestellt — beobachtet wurden. Bei Konsensusuntersuchungen ist es im übrigen durchaus möglich, ja notwendig, den unweigerlich sich einstellenden Interpretationsanteil mitzuerfassen. Jeder Mensch und zumal der von Berufs wegen dynamisch

denkende Psychoanalytiker sucht Zusammenhänge in den Phänomenen und etwas hinter ihnen. Wird die (nur angenommene, unbewußte) Strebung faktisch (= bewußt), ist sie auch anschaulich im Sinne Goethes geworden: „Das höchste wäre: zu begreifen, daß alles Faktische schon Theorie ist.“

Wegen der vielen Schwierigkeiten, denen Konsensusuntersuchungen begegnen, sind sie auch in der psychoanalytischen und psychotherapeutischen Forschung ein äußerst vernachlässigtes Stiefkind geblieben. Es gibt nur wenige empirische Veröffentlichungen zu diesem Problemkreis, deren Diskussion wir uns nun zuwenden. Unter dem leitenden Gesichtspunkt der empirischen Überprüfbarkeit der Bedingungen, die einen Konsens erleichtern bzw. erschweren, setzen sich Wallerstein und Sampson (1971) in ihrem Sammelreferat mit dem Forschungsdefizit im Bereich „Psychoanalyse und Konsensusforschung“ auseinander. Die Autoren überschreiben ihre Stellungnahme mit der kritischen Frage „The ‚consensus problem‘, or what to do when experts disagree?“ (S. 33). Das Konsensusproblem halten sie für

„die wahrscheinlich schwierigste Kernfrage, um die in der klinischen Forschung unter dem Gesichtspunkt der Begriffsbildung wie unter dem der Behandlungstechnik gerungen wird, und die in der empirisch wie in der theoretisch ausgerichteten Literatur zur klinischen Forschung unterrepräsentiert ist“.

Ausgehend von der sehr umfassenden Konsensus-Untersuchung durch Seitz (1966), auf die wir weiter unten näher eingehen werden, und in Anlehnung an Überlegungen von Strupp u. a. (1966) diskutieren Wallerstein und Sampson einige komplexe Bedingungen, welche der raschen Bildung eines Konsenses für den gegebenen Fall im Wege stehen können. Neben psychologisch-methodischen Faktoren wie Inhomogenitäten der Beobachtergruppen, Unschärfe der klinischen Phänomene und inadäquaten Beobachtungsmethoden führen sie einige klinisch-psychologische Bedingungskomplexe auf; sie sprechen „die enorme Komplexität psychologischer Daten“ (S. 35) an, die durch interpretative Akte nur unzulänglich und fehlerbehaftet verwertet werden können:

„die Tatsache, daß kognitive Funktionen unterschiedlich arbeiten und zu verschiedenartigen Problemlösungsmethoden führen, beispielsweise die verbreitete Neigung, Neues an bereits Bekanntes anzupassen, im Gegensatz zu der Tendenz, nach Neuem zu suchen und nach Gesichtspunkten, an die die anderen nicht gedacht haben“;

schließlich die sattsam bekannte Problematik des Beobachtungsfehlers,

„der in jedem Selektionsprozeß aus komplexen klinischen Datenmengen wirksam wird, eine Selektion, die durch theoretische Voreinstellungen wie auch durch emotionale blinde Flecken des Beobachters unvermeidlich beeinflußt wird“.

In diesen Ausführungen sind einige wichtige Faktoren aufgeführt, die bei klinischen Falldiskussionen das rasche Erreichen eines offenkundigen Konsenses erschweren, hinauszögern oder gar verhindern. Betrachtet man z. B. die bahnbrechende Untersuchung von Seitz (1966) zum Konsensus-Problem unter diesen kritischen Gesichtspunkten, so werden die Fallstricke, über die Seitz mit seinen Mitarbeitern gestolpert ist, klar erkenntlich. Seitz selbst formuliert rückblickend wie folgt:

„Die narzißtische Besetzung eigener interpretativer Vorstellungen erschien als ein beträchtliches (einige von uns glauben: ein abschreckendes) Hindernis auf dem Wege zu einer Verbesserung unserer Interpretationsmethode in der Psychoanalyse“ (S. 213).

Es gibt nur wenige empirische Untersuchungen zur Konsensusproblematik, die sich die von Wallerstein und Sampson (1971) skizzierte Kritik zu eigen gemacht hätten. Viel zahlreicher sind jene Arbeiten, die mit Hilfe von Beurteilerübereinstimmungs-Maßen klinischen Fragestellungen nachgehen und deren Tauglichkeit testen. Auerbach und Luborsky (1968) z. B. wollen mit der Methode der Einstufung klinischer Phänomene in Erfahrung bringen, welche Faktoren eine Therapiesitzung zu einer „good hour“ (Kris 1956) machen. Strupp und Williams (1973) interessiert mehr die Frage, inwieweit zwei Therapeuten, die mit denselben Patienten unabhängig voneinander ein Erstinterview machen, ähnliche klinische Impressionen haben, und ob sie die gleichen klinischen Phänomene als wichtige prognostische Hinweise werten würden. Ihre Untersuchung hatte dann das sogenannte „good patient cluster“ zum Resultat. Hamburg et al. (1958) beschäftigten sich in ihrer Arbeit mit der Frage, ob emotionale Phänomene genügend umschrieben sind, um verlässliche Identifizierung und Quantifizierung durch verschiedene Beobachter zu erlauben. Sie ließen die drei Affekte Angst, Ärger und Depression über eine Einwegscheibe bei solchen Patienten beobachten und auf einer Skala beurteilen, die experimentell einer Streßsituation ausgesetzt waren. Garduk und Haggard (1972) legen eine vergleichende Untersuchung der Wirkung von Interpretationen und Nichtinterpretationen auf Patienten vor. Für die Messung der Wirkung wurden entsprechende Skalen entwickelt, auf denen zwei Psychoanalytiker die Patientenantworten einstufen.

Diese Untersuchungen mögen stellvertretend für alle jene stehen, in denen zwar die Frage nach dem Konsens gestellt wird, die jedoch in erster Linie *klinische Fragestellungen* verfolgen. Es wird nach der „guten Stunde“, nach dem „guten Patienten“ gefragt, man will Verlaufsbeschreibungen machen und so weiter. Bei der skalierten Einstufung der für relevant erachteten klinischen Konzepte scheut man sich jedoch —



die unerfreulichen Folgen von Beobachtungs- und Meßfehlern sind allenthalben wohlbekannt —, nur einen einzigen kompetenten Beurteiler zu verwenden; aus der Heranziehung mehrerer Beurteiler resultiert aber die Notwendigkeit, und auch diesen methodischen Anspruch macht man sich in der klinischen Forschung erfreulicherweise zu eigen, die Übereinstimmung zwischen den Beurteilern zu überprüfen und durch ein quantitatives Maß auszudrücken. Ist eine solche Übereinstimmung unter verschiedenen Beobachtern nur hinreichend hoch — und mit ein wenig Pffiffigkeit wird sie in der Regel wohl auch hoch ausfallen —, so ist man's zufrieden und geht zum eigentlichen Thema über, ohne sich der Frage nach dem Konsens im engeren Sinne zu stellen. Im Rahmen solcher klinischen Untersuchungen bleibt es ohne Belang, auf welche Faktoren der erzielte Konsens zurückzuführen ist, beziehungsweise aus welchen Gründen die Übereinstimmung nicht noch höher ausgefallen ist.

Durch jene anderen Arbeiten, die sich durch Einhaltung quasi-experimenteller Rahmenbedingungen und die Erprobung alternativer Ansätze systematisch um eine Klärung des Konsensusproblems und damit um höhere Übereinstimmungsergebnisse bemühen, ziehen sich Überlegungen zur angemessenen definitorischen Klarheit der zu beurteilenden Konzepte wie ein roter Faden hindurch. Die Beurteilungskonzepte selbst werden hier zum Gegenstand der Forschung. Mit anderen Worten: Die Konsensusforschung sieht ihren Hauptbeitrag in erster Linie in der theoretischen Klärung der Konzepte in ihrem engen Bezug zu den klinischen Phänomenen, die im psychoanalytischen Bezugsrahmen nicht nur nach den Gesetzmäßigkeiten der multiplen Determinierung und Überdeterminierung entsprechend der „probabilistischen Natur psychischer Zustände“ (Wallerstein und Sampson, 1971, S. 35) entstehen, sondern im Akt der Beobachtung einer hypothesen- und theoriegeleiteten Wahrnehmung ausgesetzt sind. Ein wesentliches Anliegen von forschungsbezogenen Interbeobachterstudien ist es ja gerade, den Theoriegehalt manifester Äußerungen auf seine Konsensfähigkeit hin zu überprüfen. Das Forschungsinteresse gilt somit zu allererst der klinischen Verfügbarmachung der Konzepte; in den Hintergrund treten zunächst die Erforschung von Beobachtersvariablen sowie die Bereitstellung adäquater Meßtechniken.

Die theoretische Klärung psychoanalytischer Konzepte findet in neuerer Zeit Unterstützung durch eine Anzahl von Monographien, die Schlüsselbegriffe der psychoanalytischen Theorie in ihrem historischen Kontext umreißen und deren Bedeutungswandel skizzieren (vgl. die Arbeiten von Laplanche und Pontalis, 1967; Sandler et al., 1973; Nagera, 1969/70). Für die Konsensusforschung sind es insofern wichtige Vorarbeiten, als sie eine Grundlage für sachgemäße und adäquate Anwendung psy-

choanalytischer Begriffe bieten könnten und „etwas von der Mystik beseitigen, von der psychoanalytische Vorstellungen noch immer umgeben sind“ (J. Sandler et al., 1973). Andererseits können sie aber auch durch ihren einseitig historisierenden Zugang der Meinung Vorschub leisten, als seien Ansätze zur Lösung der Konsensusprobleme deshalb zum Scheitern verurteilt, weil sie notgedrungen die historischen Dimensionen theoretischer Begriffe vernachlässigen und eher die Bedeutungsvielfalt der Begriffe zugunsten ihrer phänomenalen Festlegung verkürzen.

Anhand einiger ausgewählter Untersuchungen zur Interbeobachter-Problematik wollen wir die Forschungsentwicklung und den augenblicklichen Problemstand in diesem Bereich vorstellen. Unter dem anspruchsvollen Titel „The Consensusproblem in Psychoanalytic Research“ legte Seitz (1966) eine wegweisende Studie vor. In über dreijähriger intensiver Forschungsarbeit ging er mit einem Team bekannter Psychoanalytiker der Frage nach, aus welchen Gründen Kliniker mit reicher Erfahrung und gleicher Ausbildung dasselbe psychodynamische Material unterschiedlich interpretieren.

Nur wenn die Interpretationen einer kritischen und systematischen Prüfung unterzogen werden, wird eine Beurteilung des Konsenses unter verschiedenen Beobachtern möglich sein. Dementsprechend war auch die Zielsetzung des Projekts:

„Der wesentliche Untersuchungsansatz bestand darin, daß die Mitglieder der Forschungsgruppe unabhängig voneinander Interpretationsvorschläge zu demselben therapeutischen Fallmaterial abgeben und anschließend die Gründe untersuchen, welche zu den Schwierigkeiten in der Erzielung eines Konsenses führten. Die systematische Erforschung solcher, die Übereinstimmung mindernden Einflußgrößen ermöglichte es uns, Störungen und Uneinheitlichkeiten im interpretativen Vorgehen eines jeden einzelnen zu erkennen“ (S. 210).

Diesem Plan entsprechend war es denn auch weniger eine objektivierende Untersuchung mit skalierten Beurteilungen als vielmehr der Versuch, über systematische, kontrollierte Introspektionen die interpretativen Schlußfolgerungen der einzelnen Gruppenmitglieder unter verschiedenen Gesichtspunkten zu verstehen und sie gemeinsam auf ihre Stimmigkeit mit dem Fallmaterial zu testen.

In kurzen Ausschnitten aus Behandlungsprotokollen verschiedener Patienten sowie von Stundenaufzeichnungen einer kontinuierlichen Behandlung war der „fokale Konflikt“ zu bestimmen. Dieses, ein von French übernommenes, zum Teil variiertes Behandlungskonzept, wurde in sechs Fragen ausgefächert und der Konsensusgruppe als Leitfaden für eine systematische Beantwortung vorgelegt (S. 212).

1. Worin liegt die auslösende Situation oder das auslösende Ereignis?
2. Was ist der fokale Konflikt? (Diese Frage bezieht sich auf den vorbewußten Konflikt, mit dem der Patient augenblicklich am meisten beschäftigt ist, und unterteilt sich in ein „disturbing motive“ und ein „reactive motive“.)

3. Welches sind die Widerstände gegen und die Anstrengung bei der Bearbeitung des fokalen Konflikts?
4. Welche Interpretation führte Sie zu dieser Formulierung?
5. Welcher Anhaltspunkt unterstützt Ihre Formulierung?
6. Welcher Anhaltspunkt spricht nicht für Ihre Interpretation?

Nachdem die Beurteiler unabhängig voneinander eine interpretative Aussage zum fokalen Konflikt gemacht hatten, sollte in einem zweiten Schritt jeder der Beurteiler unter Zuhilfenahme der Interpretationsversuche der anderen seine eigene erste Formulierung überarbeiten. Hierbei kam es Seitz ganz wesentlich darauf an, die Beobachtungsnähe der psychoanalytischen Konzepte zu betonen. Solche Überlegungen ließen ihn das Konzept des fokalen Konflikts von French dahingehend modifizieren, daß es eine deutliche Verankerung in beobachtbaren Prozessen erhielt. Er wies darauf hin, daß weniger die unbewußten Determinanten des Konflikts erfragt waren, als vielmehr die Vielfältigkeit dynamisch unbewußter Determinanten in einer „gemeinsamen Endstrecke vorbewußter Übertragungsträger“ (S. 214) zusammengefaßt wurde. Um es mit den Worten des Autors deutlicher zu machen:

„daß der fokale Konflikt in der aktuellen Übertragung zum Analytiker besteht und theoretisch am besten unter Berücksichtigung der Dynamik von Tagesresten verstanden werden kann“.

Auf der anderen Seite hält sich der oben aufgeführte „Fragenkatalog“ ganz im Rahmen traditioneller klinischer Fallbesprechungen. Er greift technische Begriffe wie „Abwehr“ aus der psychoanalytischen Terminologie auf, ohne deren spezifische Äußerungsform näher zu bezeichnen. Eine solche wäre in diesem Kontext etwa: Was wehrt der Patient womit ab? Von einer *Konzept*definition wird man erst dann sprechen können, wenn Verdeutlichungen dieser Art dem Beurteiler einen phänomenorientierten Anhalt geben.

Solche Beurteilungshilfen schlugen sich hingegen nicht in einer nennenswerten Interbeobachter-Übereinstimmung nieder. Die erfahrenen Kliniker ließen ihr Verständnis des fokalen Konflikts auf jeweils sehr verschiedenen Ebenen wirksam werden. Das Konzept war phänomenal zu wenig festgelegt, um in den Beurteilern ähnliche oder gar identische Interpretationsvoraussetzungen hervorzurufen.

Die Studie hat ihren Wert in der illusionslosen Offenbarung divergenter Interpretationsstrategien. Sie demonstriert mit Nachdruck, daß es in der Tat ein Konsensusproblem gibt, welches nach Meinung von Seitz vorwiegend in der Unangemessenheit der Interpretationsmethoden begründet liegt. Unterstrichen wird dieses Fazit noch durch eine weitere

Beobachtung des Autors. Anhand von Stundenzusammenfassungen sollten die Gruppenmitglieder sich auf die überzeugendste Interpretation des fokalen Konflikts einigen. War es in der Anfangsphase des Projekts noch so, daß jeder den anderen mit besonderem Engagement von der Richtigkeit seines eigenen interpretativen Vorschlags zu überzeugen versuchte, so erlahmte das Interesse an solchen Diskussionen in dem Moment, als man dazu übergang, das „agreement“ der Gruppe zu systematisieren. Wenn zwei Gruppenmitglieder jeweils ihre Interpretationen daraufhin überprüften, ob sie übereinstimmten oder nicht, funktionierte der Rest der Gruppe gewissermaßen als Schiedsrichter und nahm Stellung zu diesen Diskussionen. Dadurch wurde es immer schwieriger, einen Gruppenkonsens zu erreichen, und Resignation machte sich breit. So ideenreich das Projekt in der Formulierung psychodynamischer Konfliktkonfigurationen gewesen war, so mangelte es doch an methodenkritischen Überlegungen zum Aufbau von Untersuchungen darüber, wie die Vielfalt interpretativer Aussagen zum selben Fallmaterial zu einem empirisch gesicherten Konsens geführt werden kann.

Eine methodische Alternative zum Vorgehen von Seitz stellt das Untersuchungskonzept von Bellak und Smith (1956) dar. Sehr früh in der empirisch-psychoanalytischen Forschung griffen sie auf Schätzskalet zurück, mit deren Hilfe einfache Konzepte beurteilt werden sollten; auf diese Weise läßt sich die Güte der Beurteiler-Übereinstimmungen in einem quantitativen Maß ausdrücken. Statt wie Seitz globale Konfliktkonfigurationen beschreiben zu lassen, ließen die Autoren Patientenvariablen wie Übertragung, Einsicht, Widerstand, Schuld, Depression usw. und Abwehrmechanismen wie Projektion, Isolation, Verleugnung usw. auf solchen Schätzskalet hinsichtlich der Deutlichkeit des Auftretens einstufen. Verbatimprotokolle zweier längerer psychoanalytischer Behandlungen sowie die Bemerkungen des behandelnden Psychoanalytikers nach den Behandlungsstunden stellten die Grundlage für solche skalierten Beurteilungen dar.

In einem recht komplexen Design beurteilten jeweils zwei Psychoanalytiker mit Hilfe der Schätzskalet aus der Kenntnis der transkribierten Behandlungsprotokolle der eben abgelaufenen Woche diejenigen Variablen, die in der folgenden Behandlungsstunde, in der folgenden Woche wie auch in den nächsten vier Wochen für die jeweilige Psychoanalyse von Bedeutung sein würden. Zwei andere Psychoanalytiker gaben eine Woche später ohne Kenntnis dieser „Vorhersagen“ ihr Urteil dazu ab, welche Variablen in der vom ersten Beurteilerpaar vorhergesagten Woche für die Behandlung relevant waren und in welchem Ausmaß das

der Fall war. Neben diesen quantitativen Ratings hielten die Beurteiler ihre wesentlichen Eindrücke wie auch ihre Meinung zur weiteren Entwicklung der Behandlung in freier, nicht-skaliert Form fest. Die Paarbildungen der Beurteiler wurden in einem vierwöchigen Turnus systematisch variiert, so daß jeder Psychoanalytiker Gelegenheit hatte, mit jedem der anderen gemeinsam sowohl „Vorhersagen“ abzugeben wie auch die jeweils vergangene Behandlungswoche zu beurteilen.

Insgesamt gelangten die Beurteiler zu unterschiedlichen Übereinstimmungsergebnissen, die als tetrachorische Korrelationskoeffizienten numerisch zwischen .07 und .96 schwanken. Übereinstimmungskoeffizienten in dieser Höhe und auch in diesem Schwankungsbereich ergaben sich gleichermaßen zwischen den vorhersagenden Beurteilerpaaren wie auch zwischen jenen, die den aktuellen Stand des Behandlungsgeschehens einschätzten. Keine zufriedenstellenden Übereinstimmungen konnten hinsichtlich der Beurteilung der weiteren Entwicklung der Behandlung hergestellt werden. Kritisch ist hierzu zu bemerken, daß in diese Berechnungen nur diejenigen Variablen einbezogen wurden, die von den beiden Beurteilern gemeinsam jeweils als relevant erachtet wurden, denn nur für diese wurde ja die eingeschätzte Stärke der Bedeutung für die Behandlung auf der Skala quantitativ angegeben. Durch den Umstand, daß die gewissermaßen qualitative Entscheidung, ob eine Variable für den in Frage stehenden Behandlungsabschnitt überhaupt relevant ist oder nicht, und der hiermit zusammenhängende und gelegentlich gewiß auch aufgetretene Dissens nicht berücksichtigt wurde, erscheinen die berichteten Interbeurteiler-Übereinstimmungswerte als ungerechtfertigt hoch. Weiterhin erscheint der Versuch, mit einem Ansatz wie dem von Bellak und Smith prädiktive Aussagen über einen zukünftigen Zustand und über Veränderungen eines in psychoanalytischer Behandlung befindlichen Patienten zu machen, als sehr problematisch. Es lassen sich zweifellos zwar allgemeine, globale Entwicklungslinien einer Behandlung mit einer gewissen Zuverlässigkeit vorhersagen; aber klinisch relevante Veränderungen zu beschreiben, die in der nächsten Woche bei dem Patienten beobachtbar sein werden, ist wohl ein recht unrealistisches Unterfangen. Ein solches Vorgehen setzte ein sehr viel stärker systematisiertes und axiomatisiertes psychoanalytisches Theoriegebäude voraus (vgl. dazu Strupp, 1966). Denn es kann nicht erwartet werden, daß einem voraussagenden Psychoanalytiker alle jene Bedingungen bekannt sind, die in der folgenden Behandlungswoche die psychoanalytische Situation bestimmen werden. Empirische Regelmäßigkeiten im Auftreten von Veränderungen sind gebunden an behand-

lungstechnische Bedingungen, die in den folgenden Tagen „zufällig“ z. B. deshalb fehlen können, weil ein äußeres Ereignis eingetreten ist, das neue Randbedingungen geschaffen hat. Auf jeden Fall müßten zunächst eine große Zahl von „Wenns“ spezifiziert werden, an deren Eintreffen Änderungen gebunden sind. Wenn beim Patienten A die Disposition X (Wiederholungszwang in Form eines passiv-masochistischen Verhaltens) am Tag Y vorliegen sollte, und wenn der Analytiker B diese Widerstandsdeutungen geben sollte, sind Veränderungen aus der Reihe Z zu erwarten. Mit diesem Beispiel können wir natürlich nur auf einige Probleme von Voraussagestudien hinweisen, um anzudeuten, daß vor Konsensus-Untersuchungen eine umfangreiche klinische Systematik geleistet werden müßte, um von „Prophezeiungen“ zu „Vorhersagen“ und damit zu einem Konsens kommen zu können.

Trotz dieser Einwände muß man mit den Autoren feststellen, daß ihre Untersuchung die Notwendigkeit und die Möglichkeit der Durchführung und auch der klinisch sinnvollen Bewertung von empirischer, quantitativer Konsensusforschung deutlich macht und mit insgesamt eindrucksvollen empirischen Befunden belegt.

Erst ein Jahrzehnt später kam Strupp (1966) auf die Untersuchung von Bellak und Smith zurück und führte die Überlegungen zur Interbeurteiler-Problematik fort. Vornehmlich unter methodenkritischen Gesichtspunkten diskutiert er erneut die Möglichkeit, in der klinischen Forschung zu höheren Übereinstimmungen zu gelangen. Seine experimentellen Untersuchungen gelten in erster Linie dem Versuch, zwischen klinischen Konzepten hinsichtlich ihres Inferenzgrades und ihrer Konsensfähigkeit zu differenzieren. Hierfür stellte Strupp in Anlehnung an Bellak und Smith 35 Prozeßvariablen zusammen, und zwar nicht nur solche, die auf eine Beschreibung des Patienten gerichtet sind, sondern ebenso Variablen, die das Therapeutenverhalten charakterisieren. Diese Variablen wurden nach dem Grad der Inferenz, d. h. nach dem Grad der Beobachtungsnähe bzw. Beobachtungsferne auf einer siebenstufigen Skala mit umschriebenen Skalenpunkten (z. B. von „almost directly observable“ bis „almost impossible to observe directly“) beurteilt. Beurteiler waren elf klinische Mitarbeiter. Als relativ beobachtungsfern erwiesen sich in dieser Untersuchung Konzepte wie die orale, anale und phallische Strebung; „elation“ hingegen ist nach Meinung der Beurteiler ein ziemlich beobachtungsnahes Konzept.

In einem zweiten Teilerperiment beobachteten vier klinisch relativ unerfahrene Beurteiler elf therapeutische Sitzungen über Einwegscheibe und Lautsprecheranlage und schätzten die Prozeßvariablen anhand von

Skalen hinsichtlich der Intensität ein. Als Maß für die Höhe der Übereinstimmung zwischen den Beurteilern über die elf Sitzungen hinweg wurde der Intraklassen-Koeffizient (Haggard, 1958) berechnet. Die Übereinstimmungswerte für die einzelnen Variablen schwankten zwischen .08 und .58.

In einem dritten Schritt wurde nach der korrelativen Beziehung zwischen Beobachtungsnähe bzw. Beobachtungsferne der Variablen und der Höhe der Interbeurteiler-Übereinstimmung gefragt. Es ergab sich ein signifikanter Rangkorrelationskoeffizient von  $r = .71$ , was bedeutet, daß die Interbeurteilerübereinstimmung in der Regel um so höher ausfällt, je deutlicher das beurteilte klinische Konzept auf beobachtbare Prozesse bezogen ist. Umgekehrt gilt: je häufiger oder intensiver bei der Beurteilung eines klinischen Konzepts auf dezidiertes Theoriewissen zurückgegriffen werden muß, desto niedriger fällt in der Regel die Übereinstimmung zwischen Beurteilern aus. So bestätigt sich, was für den Kliniker evident ist: „daß eine hauptsächliche Determinante der Interbeobachter-Übereinstimmung der Grad an Inferenz ist, der für ein klinisches Urteil erforderlich ist“ (Strupp, 1966, S. 448).

Die Resultate von Strupp verlieren allerdings dadurch an Überzeugungskraft, daß beim ersten Experiment klinisch ausgebildete Beurteiler, beim zweiten hingegen klinisch relativ unerfahrene Beurteiler herangezogen wurden. (Auf diesen Umstand mag es zurückzuführen sein, daß sich für ein so stark im Erleben und in klar beobachtbaren Ausdrucksbewegungen wurzelnde Konzept wie „elation“ — das im ersten Experiment als am beobachtungsnahesten beurteilt wurde — durchschnittlich eine negative Interbeobachterübereinstimmung ergab.) Dennoch drängt sich aus diesen Experimenten eine Schlußfolgerung auf: Um auch klinische Konzepte mit einem hohen Inferenzgrad und theoretische Begriffe konsensfähig zu machen, ist eine empirienähe Umformulierung in beobachtungsnahe Definitionen der Konzepte unerlässlich. Bei unseren eigenen empirischen Untersuchungen haben wir diesen Gedanken aufgegriffen und konsequent verfolgt. Hierbei ist man jedoch der Versuchung ausgesetzt, die Konzepte zu eng zu definieren und damit ihren Bedeutungsraum innerhalb von Theorie und Praxis über die Maßen einzuengen. Bernfeld (1932) hat die Psychoanalyse einmal eine „Spurenwissenschaft“ genannt. Es ist wesentlich, daß ein Psychoanalytiker mit Hilfe z. B. des Begriffs der Kastrationsangst auch die feinsten Spuren derselben im Denken und Verhalten eines Patienten findet und nicht erst solche Ausprägungen, die den Begriff in reiner Form erfüllen, wobei dann eine hohe Interbeobachter-Übereinstimmung

zu erwarten wäre. Auf der anderen Seite muß man fragen, wie gut ein Konsens zwischen einem Psychoanalytiker und seinem Patienten über Deutungen begründet sein mag, wenn darüber keine konsensuelle Validierung zwischen Experten zu erreichen ist.

Lehrreich sind in diesem Zusammenhang die Arbeiten von Meyer et al. (1969) und von Luborsky et al. (1973), die einige ungelöste Fragen der Konsensusproblematik thematisieren. Meyer et al. entwickelten für die Konstruktion eines mehrdimensionalen Persönlichkeitsfragebogens auf der Grundlage der psychoanalytischen Charakterologie (vgl. dazu Kline, 1972) ein System von Kategorien, das in globaler Betrachtung zunächst die psychoanalytische Phasenlehre (orale, anale, phallische und zusätzlich phasenunspezifische Merkmale) widerspiegelt. Jede dieser Globalkategorien war in mehrere „spezifische“ Einzelskalen unterteilt, z. B. der Bereich „Oralität“ aufgegliedert in „zärtliche Impulse“, „oral-aggressive Impulse“, „Urvertrauen“ etc. In dieses Kategoriensystem wurden von drei Psychoanalytikern, die unabhängig voneinander vorgehen, zum einen 159 globale Eigenschaften der psychoanalytischen Charaktere, wie sie in der Fachliteratur beschrieben sind, eingeordnet, zum anderen 687 Verhaltensbeschreibungen und Einstellungen zu konkreten Problemen, die ebenfalls der Literatur entnommen wurden und nicht für „psychoanalytisch irrelevant“ (S. 832) erachtet worden waren. Hinsichtlich der globalen Eigenschaften ergab sich, daß die drei Beurteiler untereinander zwar recht gut übereinstimmten (89 % dieser globalen Items wurden von mindestens zwei der drei Psychoanalytiker gleich beurteilt), daß die Übereinstimmung mit den in der Literatur genannten Zuordnungen jedoch unbefriedigend niedrig ausfiel (56 Prozent). Unter Berücksichtigung der Mehrdeutigkeit von Charaktereigenschaften wie z. B. „eifersüchtig“ interpretieren die Autoren diesen Befund: „Unsere Items erschienen daher in bezug auf ihre Klassifizierbarkeit nicht etwa als zu leicht, wie wir anfangs befürchtet hatten, sondern als zu schwer, weil zu abstrakt“ (S. 830.) Dieser Überlegung entspricht auch das Übereinstimmungsergebnis bei den spezifischen Verhaltensbeschreibungen, das noch weniger günstig als bei den globalen Eigenschaften ausfiel; es ist also sicherlich zu naiv, anzunehmen, daß eine Annäherung des zu beurteilenden Materials an direkt beobachtbares Verhalten oder an äußere Sachverhalte *zwangsläufig* eine höhere Interbeurteiler-Übereinstimmung mit sich bringt. Es muß stets zugleich im Auge behalten werden, welche Konzepte und Begriffe zur Beurteilung des Materials anstehen; erst wenn das Beurteilungsmaterial auf die Beurteilungskonzepte abgestimmt wird, kann ein befriedi-



gender Konsens erwartet werden. Eine einfache Verhaltensbeschreibung wie „Ich liege gern lange im warmen Badewasser“ ist im Hinblick auf die phasenspezifischen Kategorien von Meyer et al. immer noch zu komplex und vor allem zu vieldeutig, so daß eine übereinstimmende Zuordnung nicht zustande kommen kann.

Im Zusammenhang mit der Frage nach dem Komplexitätsgrad des Beurteilungsmaterials möchten wir auf die Beziehung zu dessen Umfang, wenigstens sofern es sich um Ausschnitte einer psychoanalytischen oder psychotherapeutischen Behandlung handelt, hinweisen. Es wird ja nach der optimalen Länge eines Textes gesucht, die zu einer maximalen Beurteilerübereinstimmung führt. Eine ganze Behandlungsstunde von ca. 50 Minuten wird oft als zu lang angesehen, eine transkribierte Textzeile oder ein Satz als zu kurz. Wir bezweifeln allerdings, ob eine solchermaßen formulierte Frage überhaupt allgemein sinnvoll beantwortet werden kann; es scheint doch eher so zu sein, daß die Frage nach dem optimalen Umfang des Beurteilungsmaterials nur im Zusammenhang mit dem zu beurteilenden klinischen Konzept gesehen werden kann. Eine solche Überlegung scheint uns einige Widersprüchlichkeiten in der Interpretierung niedriger Interbeurteiler-Übereinstimmungen aufzuklären, so beispielsweise bei Luborsky et al. (1973), die feststellen,

„daß die Fünfminuten-Ausschnitte allzu unzureichende Hinweise zu enthalten scheinen, um ein so komplexes Konzept (wie ‚Übertragung‘, Verf.) zu beurteilen, denn Analytiker sind gewohnt, ihr Urteil auf der Grundlage von mehreren Sitzungen abzugeben“ (S. 73). Wenige Zeilen später allerdings führen sie das Resultat, daß das Konzept „Gesamtübertragung“ (global transference) weniger deutlich übereinstimmend beurteilt worden war als die „objektgebundene Übertragung“, auf die Schwierigkeit zurück, „einen Konsens zu erhalten, wenn jeder der Beurteiler womöglich unterschiedliche Teilaspekte aus den Fünfminuten-Ausschnitten berücksichtigt“.

Einige Konzepte erfordern vielleicht eine längere Äußerungssequenz, um sinnvoll beurteilt werden zu können; sie mag freilich für andere Konzepte eine verwirrende Vielfalt an mehrdeutig interpretierbaren Informationen beinhalten (vgl. hierzu auch Mintz und Luborsky, 1971).

Wir wollen zum Abschluß dieser Literaturübersicht ein Fazit ziehen und, auf Strupp (1966) und Wallerstein und Sampson (1971) gestützt, einige Bedingungskomplexe aufführen, die bei empirischen Untersuchungen die Konsensbildung erschweren bzw. bei angemessener Berücksichtigung überhaupt erst einen nennenswerten Konsens ermöglichen:

1. Wie sich beispielsweise bei Strupp (1966) zeigte, gab der unterschiedliche klinische Ausbildungsstand seiner Beurteiler berechtigten Anlaß zur Kritik an den Resultaten; eine Beurteilergruppe bestand aus

erfahrenen Psychoanalytikern, die andere Gruppe aus Studenten in klinischer Ausbildung. Ganz allgemein ist eine Homogenität der Beurteiler hinsichtlich wesentlicher Variablen zu fordern, so z. B. hinsichtlich der klinischen Erfahrung und der Erfahrung bzw. der Ausbildung im Umgang mit skalierten Einschätzungen („Skalentraining“). Je nach Fragestellung ist auch eine unterschiedliche theoretische Voreinstellung der erfahrenen Kliniker in Betracht zu ziehen, die sich durchaus in einem abweichenden Umgang mit den vorgegebenen Konzepten und mit dem klinischen Material niederschlagen kann.

2. Im allgemeinen erscheint es sinnvoll, statt einer umfangreichen Liste klinischer Prozeßvariablen je nach Fragestellung nur einige wenige relevante Konzepte in die Konsensus-Untersuchung einzubeziehen, diese dann allerdings durch beobachtungsnahe Umschreibungen in Form von Definitionen zu operationalisieren. Je zutreffender und präziser solche operationalen Definitionen formuliert sind (vgl. z. B. Gottschalk und Gleser, 1969), desto einheitlicher werden sie auf das vorgelegte Material angewandt werden. Mit anderen Worten: je besser die Operationalisierung (im weiteren Sinne) eines klinischen Konzepts gelingt, desto eher wird auch ein Konsens zu erreichen sein. Allerdings müssen die Beurteiler dahingehend instruiert werden, daß in den Beurteilungsvorgang möglichst nur die in den operationalen Definitionen enthaltenen Parameter einfließen sollten, keinesfalls die „privaten“ Auffassungen der Beurteiler. Inwieweit sich dieses Desideratum realisieren läßt, muß wohl zunächst eine offene Frage bleiben.

3. Die Länge der Beurteilungseinheit, der Umfang des zu beurteilenden klinischen Textmaterials sollte den verwendeten Konzepten optimal entsprechen. Gründliche theoretische klärende Vorüberlegungen können sich hier als sehr hilfreich erweisen. Eventuell müssen entsprechende Voruntersuchungen durchgeführt werden, um zu einer Annäherung an die *idealiter* optimale Textlänge im Hinblick auf gegebene klinische Konzepte zu gelangen. Unzutreffend ist jedenfalls der voreilige Schluß, daß die Verwendung von kürzeren Beurteilungseinheiten oder gar von einzelnen Worten zwangsläufig zu höheren Übereinstimmungen führt (vgl. dazu die Untersuchung von Meyer et al., 1969).

4. Um auch eine von subjektiven Eindrücken losgelöste Beurteilung der Güte der in einer empirischen Untersuchung angegebenen Übereinstimmung zu ermöglichen, ist die Angabe von quantitativ-statistischen Maßen als Ausdruck für die Güte der erzielten Übereinstimmungen uner-

läßlich. Denn die Aussage des Klinikers, „die Übereinstimmung war recht hoch“, enthält weniger Information und eignet sich weniger gut für weitere Vergleiche als die eher statistische Aussage: „Die Beurteiler stimmten zu 65 Prozent überein“ oder „Es ergab sich eine Interbeobachterübereinstimmung in Höhe von  $r = .71$ .“ Allerdings werden in der Literatur hierfür unterschiedliche statistische Methoden verwandt — die teilweise nicht einmal benannt sind —, was die Vergleichbarkeit der Werte erschwert.

Unsere eigenen, im folgenden referierten empirischen Untersuchungen basieren auf diesen Erfahrungen. Sie mögen als Beitrag zur Methodik wie auch zur klinischen Theoriebildung innerhalb der Konsensusforschung verstanden werden. Im Hinblick auf die Kontroversen bezüglich eines statistischen und eines klinischen Forschungsansatzes (Meehl, 1954) möchten wir eine Bemerkung von Holt (1958, S. 12) voranschicken, der sich gegen diese unnötige Dichotomisierung in der klinisch-psychologischen Forschung wendet:

„Statt dessen sollten wir versuchen, für jede einzelne Vorhersageuntersuchung das optimale Verhältnis zwischen den von der Statistik geforderten Methoden und dem einfühlenden Urteil des Klinikers zu schaffen. Wir müssen für das richtige Gleichgewicht zwischen Freiheit und Einengung sorgen, ein Gleichgewicht, das sich in Abhängigkeit von der Art des untersuchten Verhaltens sehr wohl zur einen oder anderen Seite hin verschieben kann. Solche ausgewogenen Verhältnisse können wir aber nur dann erreichen, wenn klinisch und statistisch orientierte Forscher streitsüchtige, konkurrierende Einstellungen aufgeben und voneinander zu lernen suchen.“

## *II. Eigene empirische Untersuchungen*

### *Angsteinstufungen anhand von Stundenprotokollen<sup>7</sup>*

In den Mittelpunkt unserer Einstufungsuntersuchungen haben wir psychoanalytische Angstkonzepte gestellt. Eine theoretisch-klinische Klärung von Angstformen sowie die Frage ihrer Konsensfähigkeit schien uns mit Freud deshalb so wichtig, weil die Angst „das Grundphänomen und Hauptproblem der Neurose“ (1926 a, S. 175) ist. Die empirischen Untersuchungen erfolgten anhand der Verbatimprotokolle eines jungen Mannes, der an einer so schweren Angstsymptomatik litt, daß er fast zwei Jahre wegen seiner frei flottierenden Ängste stationär behandelt werden mußte.

<sup>7</sup> Eine ausführlichere Darstellung und Diskussion dieser Untersuchung in monographischer Form befindet sich in Vorbereitung (Kächele, Thomä und Grünzig, 1976).

Über weite Strecken der Analyse befand sich der Patient in einer Art von Panik. So war der Patient etwa unfähig, ohne Angst auch nur eine Minute zu warten oder ohne Beunruhigung Wetter- und Temperaturveränderungen hinzunehmen. Wegen der Schwere seiner Herzängste und seiner völligen Arbeitsunfähigkeit hatte der Patient seine Berufstätigkeit unterbrechen müssen. Nach zweijähriger stationärer Behandlung war der Patient fähig, für ein weiteres Jahr ambulant fünfmal wöchentlich zur Analyse zu kommen. Seit etwa eineinhalb Jahren hat der Patient seinen Beruf wieder aufgenommen.

Die Tatsache, daß es kaum einen angstfreien Lebensbereich für den Patienten gab, kennzeichnet die Schwere der Neurose: Sein Ich war, um Freuds Ausdruck abzuwandeln, fast nur noch Angststätte (Freud 1926 a, S. 171).

Wie Beobachtungen in klinischen Fallseminaren immer wieder zeigen, kommen Übereinstimmungen zwischen Klinikern oft sehr rasch, oft aber auch gar nicht zustande. Solche Übereinstimmungsleistungen sollen im folgenden, aus der klinischen Feldsituation herausgelöst, semi-experimentell<sup>8</sup> mit Hilfe von Schätzskalen näher untersucht werden. Wie wir weiter oben ausführlicher diskutierten, ist es bei derartigen Einstufungsuntersuchungen erforderlich, die Konzepte durch beobachtungsnahe Umschreibungen in Form von Definitionen zu operationalisieren. In Anlehnung an Gottschalk und Gleser (1969) haben wir sechs Formen der Angst mit den von Gottschalk vorgeschlagenen Definitionen zum Gegenstand unserer Untersuchungen zur Konsensus-Problematik gemacht. Einerseits scheint uns nämlich die von ihm vorgenommene Auswahl den Kern der psychoanalytischen Angsttheorie zu repräsentieren, ohne die theoretische Diskussion festzulegen, andererseits ist seine Auswahl durch das Interesse am Sprachverhalten bestimmt worden, was der Erforschung des uns zur Verfügung stehenden Verbatimmaterials entgegkommt.

### *Konzeptdefinitionen*

Zur Verdeutlichung dessen, was die verschiedenen Angstkonzepte meinen und in welchem theoretischen Bezug sie stehen, diskutieren wir sie kurz und geben jeweils am Ende die operationale Definition an:

#### *a) Trennungsangst*

Die Trennungsangst gehört zur präödiptalen Entwicklungsphase, also zur Zwei-Personen-Beziehung. Der Verlust des Objektes, d. h. der Mutter, läßt das erlebende Kind tatsächlich „mutterseelenallein“ in Panik zurück. In dieser Nähe zur symbiotischen Einheit können Objektverlust

<sup>8</sup> Shakow (1960) spricht von einem semi-naturalistischen Zugang.

und Ich- bzw. Selbstverlust als identisch erlebt werden (vgl. dazu M. Mahler, 1958; K. Angel, 1972). Somit kann die Gefahrensituation des Objektverlustes als Paradigma traumatischer Hilflosigkeit bezeichnet werden.

Bei neurotischen Trennungsängsten findet man regelmäßig unbewußte Aggressionen, welche die Furcht, das Objekt zu verlieren, überhaupt erst hervorrufen. Man befürchtet, das Objekt zerstört zu haben, woraus sich viele Verhaltensweisen von Phobikern erklären, die sich der Anwesenheit und Integrität des (unbewußt zerstörten) Objekts versichern müssen.

Eine handhabbare Kurzdefinition von Trennungsangst kann nicht in der Lage sein, diese skizzierten psychogenetischen Zusammenhänge vollständig wiederzugeben. Die Zusammenstellung von Gottschalk scheint uns dennoch recht umfassend: Hinweise auf Verlassen, Verlassenheit, Einsamkeit, Verstoßung, Verlust von Unterstützung, von Liebe oder Liebesobjekt oder eine damit verbundene Bedrohung.

#### *b) Kastrationsangst*

Die universelle Gültigkeit der Kastrationsangst leitet Freud aus der biologischen Konstitution ab. Der anatomische Geschlechtsunterschied bedingt, daß der kleine Junge beim Anblick eines weiblichen Körpers dessen Penislosigkeit entdeckt und sieht, daß das von ihm so hoch geschätzte Glied nicht notwendig mit zum Körper gehört (Freud, 1938). Die Erlebniskategorie „Trennung“ (und die damit verbundenen Ängste) erhält nun einen entwicklungspsychologisch neuen Inhalt, nämlich die Kastrationsangst. Erscheint die Trennungsangst als Reaktion auf das Vermissen des Objekts, sieht Freud die Analogie darin, „daß auch die Kastrationsangst die Trennung von einem hochgeschätzten Objekt zum Inhalt hat“ (1926 a, S. 167). Somit bezieht sich das Konzept „Kastrationsangst“ nicht nur auf seine eingeeengte anatomische Bedeutung, auf den Verlust des Penis oder des ganzen Geschlechtsteils, sondern auch auf deren symbolische Äquivalente. Auch Drohungen, die sich nicht im engeren Sinn auf das Abschneiden des Genitals richten, können so erlebt werden: Bestrafungen, die sich auf das Spielen am Geschlechtsteil beziehen, werden zur „Kastrationsdrohung“. Allgemein können körperliche Verletzungen oder seelische Kränkungen ebenso als Kastrationsdrohung erlebt werden. Diese Kurzbeschreibung enthält wesentliche Elemente der Definition von Gottschalk: Hinweise auf Verletzung, physische Beschädigung oder Angst vor Verletzung, weiterhin

Hinweise, die eine Bedrohung der Potenz und des Könnens überhaupt repräsentieren.

### *c) Schuldangst*

Die Fähigkeit, Angst vor dem Über-Ich zu entwickeln, kennzeichnet den Schritt, der die Überwindung der ödipalen Dreiersituation einleitet:

„Mit dem Unpersönlichwerden der Elterninstanz, von der man die Kastration befürchtete, wird die Gefahr unbestimmter. Die Kastrationsangst entwickelt sich zur Gewissensangst, zur sozialen Angst. Es ist jetzt nicht mehr so leicht anzugeben, was die Angst befürchtet. (...) Allgemeiner ausgedrückt, ist es der Zorn, die Strafe des Über-Ichs, der Liebesverlust von dessen Seite, den das Ich als Gefahr wertet und mit dem Angstsignal beantwortet“ (Freud, 1926 a, S. 170).

Mit dieser Bestimmung bleiben freilich solche Schuldängste ausgeklammert, die sich auf die Entwicklung des präödipalen „Gewissens“ beziehen (vgl. dazu A. H. Modell, 1975). Bezeichnungen wie Über-Ich-Angst, Gewissensangst, soziale Angst und Schuldangst werden weitgehend synonym verwandt. Die formalisierte Beurteilung stützt sich auf folgende kurze Definitionen: Hinweise auf Kritik, moralische Verurteilung, Verdammung, Verurteilung, auf Schuld oder Bedrohung durch Schuld.

### *d) Beschämungsangst*

Die Entwicklung der Ich-Psychologie und die jüngst aufgekommene Diskussion über den Narzißmus und das Selbst richten die Aufmerksamkeit auf jene Vorgänge, die sich im Ich als Angststätte abspielen. Die Inhalte der Gefahrensituation und ihre Zuordnung zur Triebentwicklung (Objektverlust, Liebesverlust, Kastration) werden nunmehr daraufhin betrachtet, welche Konsequenzen im Ich bzw. Selbst entstehen, wenn es sich hilflos fühlt. Das Selbst stellt man sich hierbei als eine Instanz vor, die der Ich-Wahrnehmung dient und als solche in kritische Distanz zum Angstaffekt treten kann (vgl. dazu Sandler und Joffe, 1967). Die Hilflosigkeit erhält dadurch weitere Dimensionen, die zum Beispiel als Angst vor Kontrollverlust, vor Verlust der Autonomie oder als narzißtische Kränkung bezeichnet werden. Es erscheint sinnvoll, diese Form antizipatorischer Wahrnehmung als eine eigene Angstform anzusetzen, die in ganz besonderer Weise den narzißtischen Aspekt des Angsterlebens abbildet.

1008 *H. Thomä und andere*

Für die Verwendung des narzißtischen Gesichtspunktes in dieser Untersuchung war allerdings eine Vereinfachung vorzunehmen. Im Mittelpunkt der kränkenden Erfahrung steht die Selbstwertproblematik, die sich phänomenologisch eng an das Schamgefühl anschließen läßt. Deshalb halten wir die von Gottschalk et al. (1969) in Anlehnung an Piers und Singer (1953) vorgeschlagene Kurzdefinition für vertretbar: Hinweise auf das Gefühl der Lächerlichkeit, der Kränkung, des Ungnügens, der Scham und Erniedrigung.

#### *e) Todesangst*

„Als letzte Wandlung dieser Angst (Kastrationsangst und Schuldangst, Verf.) vor dem Über-Ich ist mir die Todes- (Lebens-)Angst, die Angst vor Projektion des Über-Ichs in den Schicksalsmächten erschienen“ (Freud, 1926 a, S. 170). Im Gegensatz zu den anderen neurotischen Ängsten geht die Todesangst nicht auf eine vergangene Situation zurück, sondern antizipiert etwas Zukünftiges. Dementsprechend hält Freud daran fest, daß der Todesangst im Unbewußten nichts entspricht, was dem Begriff der Lebensvernichtung Inhalt geben könnte, denn „etwas dem Tod Ähnliches ist (...) nie erlebt worden“ (1926 a, S. 160). Es ist anzunehmen, daß auch der Realangst des Todes, die als zukünftiges Ereignis auf uns zukommt, ein infantiles Muster des Verlassenwerdens oder der Trennung entgegenkommt (vgl. dazu Stern, 1972). Soweit die Todesangst als neurotische Angst auftritt, ist sie einerseits als Verarbeitung des Kastrationskomplexes, andererseits als Ausdruck der Trennungsängste der prägenitalen Periode zu verstehen.

Die von uns verwendete Kurzdefinition bleibt dicht am manifesten Gehalt: Hinweise auf Tod, Sterben, Todesdrohung oder Angst vor dem Tod, bezogen auf sich selbst oder andere.

#### *f) Diffuse Angst*

Hierunter fällt die von Freud „frei flottierende Angst“ genannte Angstform. Der Angstaffekt hat sich gewissermaßen von der spezifischen Gefahrensituation losgelöst und dann verselbständigt; zumindest ist seine Herkunft unklar. Der Zustand kann auch als ein Darniederliegen der Abwehrvorgänge beschrieben werden, deren Ziel, die Angst durch Symptombildung zu binden, nicht erreicht wurde.

So sollen mit dieser Beurteilungsdimension besonders auch jene somatischen Angstäquivalente erfaßt werden, die als „objektlose“ Ängste

nicht sicher einer der obigen Gefahrensituationen zugeordnet werden können. Für die Beurteiler wurde folgende Kurzdefinition vorgegeben: Hiermit ist der objektlose Zustand der frei flottierenden Angst gemeint, die Überschwemmung durch Angst mit besonderer Akzentuierung der Somatisierung.

In einer Tabelle fassen wir die Angstformen und ihre Operationalisierungen zusammen:

Tabelle 1

<i>Beurteilungskonzepte</i>	<i>Definitionen</i>
1. Trennungsangst	Hinweise auf Verlassen, Verlassenheit, Einsamkeit, Verstoßung, Verlust von Unterstützung, von Liebe oder Liebesobjekt oder eine damit verbundene Bedrohung.
2. Kastrationsangst	Hinweise auf Verletzung, physische Beschädigung oder Angst vor Verletzung, weiterhin Hinweise, die eine Bedrohung der Potenz und des Könnens überhaupt repräsentieren.
3. Schuldangst	Hinweise auf Kritik, moralische Verurteilung, Verdammung, Verurteilung, auf Schuld oder Bedrohung durch Schuld.
4. Beschämungsangst	Hinweise auf das Gefühl der Lächerlichkeit, der Kränkung, des Ungenügens, der Scham und Erniedrigung.
5. Todesangst	Hinweise auf Tod, Sterben, Todesdrohung oder Angst vor dem Tod, bezogen auf sich selbst oder andere.
6. Diffuse Angst	Hiermit ist der objektlose Zustand der frei flottierenden Angst gemeint, die Überschwemmung durch Angst mit besonderer Akzentuierung der Somatisierung.

Aus der erwähnten psychoanalytischen Behandlung wählten wir aus dem Behandlungszeitraum von ca. 500 Stunden eine Stichprobe von 55 Behandlungsstunden aus, die sich gleichmäßig über die Behandlung verteilen und als repräsentativ angesehen werden können. Die transkribierten Verbatimprotokolle dieser Stichprobe, die pro Stunde etwa 10–15 maschinengeschriebene Seiten umfassen, stellen das Material für die Beurteilung der Angstkonzepte dar.

Die Verwendung von Schätzskaleten für die Beurteilung klinischer Konzepte — auch in der psychoanalytischen Forschung mittlerweile zum Standard-Methodenrepertoire gehörig — bringt neben anderem den Vorzug mit sich, daß sich die Übereinstimmung zwischen Beurteilern in einem quantitativen Maß ausdrücken läßt und damit eine eher objektive Bewertung ermöglicht.

Bei einer subjektiven Beurteilung des Grades an Übereinstimmung durch mehrere Beobachter (wie sie beispielsweise in einem klinischen Fallseminar eintritt) stellt sich die Frage, inwieweit nun die Beobachter in ihrer Bewertung der Übereinstimmungen innerhalb des Fallseminars einig sind. Diese Frage nach der Inter-Beobachter-Über-



1010 *H. Thomä und andere*

einstimmung läßt sich nur durch eine weitere Gruppe von Beurteilern beantworten, die nun ihrerseits die Beurteiler erster Ordnung hinsichtlich ihrer Übereinstimmungsleistungen beurteilen. Nun kann aber ebenso der Grad der Verständigung unter den Beurteilern zweiter Ordnung hinterfragt werden, was wiederum eine weitere Beurteilergruppe erforderlich macht. Dieser Prozeß läßt sich ad infinitum fortsetzen, ohne daß eine zufriedenstellende Beantwortung der Ausgangsfrage — wie stimmen die Teilnehmer am Fallseminar überein? — gefunden werden könnte.

Für die Einstufungen der Angstkonzepte verwandten wir zum einen eine Skala, auf der die Intensität der Angstthematik, d. h. ihre dynamische Wirksamkeit eingeschätzt werden soll. Zunächst soll angegeben werden, ob das jeweilige Konzept überhaupt wirksam ist, und wenn ja, in welchem Ausmaß das der Fall ist.

Tabelle 2

Das jeweilige Angstkonzept ist dynamisch			
		wirksam	oder nicht wirksam
		0	0
wenig	mittel	stark	
0	0	0	
1	2	3	

Da zwischen bewußten und unbewußten Gefahrensituationen unterschieden werden kann, wurde eine weitere Skala „Grad der Bewußtheit“ eingeführt, auf der im Sinne des topographischen Modells über den Ort der Angst beim Patienten entschieden wird. Diese „Bewußtheitsskala“ wurde nur dann benutzt, wenn das jeweilige Konzept als dynamisch wirksam angesehen wurde.

unbewußt	0	—	0	—	0	—	0	—	0	—	0	—	0	—	0	bewußt
	1		2		3		4		5		6		7			

Es wurde also jede der 55 ausgewählten transkribierten Behandlungsstunden unter Verwendung der beiden Skalen hinsichtlich jedes der sechs Angstkonzepte beurteilt, und zwar von dem behandelnden Psychoanalytiker (A), einem Arzt in psychoanalytischer Ausbildung (B) und einer klinischen Psychologin (C).

*Ergebnisse*

Um die statistischen Prozeduren für die Bestimmung der Inter-Beurteiler-Übereinstimmung zu vereinfachen, haben wir die „Intensitäts-“ und die „Bewußtheits-Skala“ auf einer einzigen, gemeinsamen Skala abge-

bildet. Hierbei wurde der Überlegung Rechnung getragen, daß die Einschätzung der dynamischen Wirksamkeit eines Konzepts der Beurteilung der Bewußtheit logisch und bedeutungsmäßig vorgeordnet ist. Die folgende Tabelle enthält das Ausmaß der Inter-Beurteiler-Übereinstimmungen in Form von Produkt-Moment-Korrelationskoeffizienten zwischen den jeweiligen Beurteilerpaaren wie auch für die Beurteilergruppe insgesamt.

(Das Todesangstkonzept bleibt hier und in der weiteren Darstellung unberücksichtigt, da dieses Konzept übereinstimmend zu häufig als dynamisch unwirksam angesehen wurde.)

Tabelle 3

*Inter-Beurteiler-Übereinstimmung \*\* (Produkt-Moment-Korrelation)*

	A/B	A/C	B/C	r	Strupp (1966)	Strupp (1973)	Auerbach (1968)
Katastrationsangst	.31	.38	.29	.33	—	—	—
Trennungsangst	.55	.53	.30	.46	—	—	—
Schuldangst	.37	.34	.31	.34	—	—	—
Beschämungsangst	.29	.04	.07	.13	—	—	—
Diff. Angst	.59	.53	.49	.54	—	—	—
Anxiety	.42 *	.36 *	.29 *	.36 *	.38	.22	.50
Hostility	—	—	—	—	.36	.46	.57
Dominance	—	—	—	—	.34	.35	—
Depression	—	—	—	—	.50	—	.52
Dependence	—	—	—	—	.58	—	.37
Empathy	—	—	—	—	.32	—	.50
Defensiveness	—	—	—	—	.08	.46	—

N = 55: P = .05 bei r = .27  
P = .01 bei r = .35 (zweiseitig)

Die Homogenitätsprüfung der Koeffizienten der drei Beurteilerpaare ergab in keinem Fall eine bedeutsame Abweichung.

\* Diese Werte stellen die spaltenweise gemittelten Korrelationskoeffizienten dar.

\*\* Zum Vergleich sind die publizierten Inter-Rater-Übereinstimmungswerte anderer Autoren (Strupp et al., 1966; Strupp, 1973; Auerbach et al., 1968) herangezogen worden.

Auffallend ist die durchgängig geringe Übereinstimmung in bezug auf das Konzept „Beschämungsangst“, was auf die theoretisch wenig zufriedenstellende Begründung des Konzepts hinweist. Die übrigen Inter-Beurteiler-Koeffizienten liegen völlig im Rahmen der Angaben der anderen Autoren. Führt man sich aber nun die Aussagekraft solcher Korrelationskoeffizienten — selbst bei zufallskritischer Absicherung — deutlicher vor Augen, so bedeutet die maximale Übereinstimmung

in Höhe von  $r = .59$  (Diffuse Angst zwischen Beurteiler A und B), daß diese beiden Beurteiler nur zu etwa 36 % ein gemeinsames Beurteilungsverhalten aufweisen. Mit anderen Worten: der Grad der Nicht-Übereinstimmung ist immer noch höher als der Grad der Übereinstimmung; erst bei einer Inter-Beurteiler-Korrelation von  $r = .71$  sind beide Anteile gleich stark ausgeprägt. Somit können die hier vorgelegten Resultate schwerlich als solide Grundlage beispielsweise für nachfolgende Untersuchungen zur Konstruktvalidierung herangezogen werden. Auch Meyer et al. (1969) haben die Konstruktion eines mehrdimensionalen Persönlichkeitsfragebogens auf psychoanalytischer Grundlage vorerst zurückstellen müssen<sup>9</sup>, da die Interbeobachter-Übereinstimmung in bezug auf die Zuordnung von Items (Verhaltensbeschreibungen sowie Eigenschaften, die sich aus der psychoanalytischen Charakterlehre ableiten lassen) zu Kategorien nicht die geforderte Höhe von .70 bis .80 erreichte.

### *Weiterführende Überlegungen*

Geht man nun der Frage nach, worauf sich die mittlere Höhe der Interrater-Übereinstimmungen zurückführen läßt, so bieten sich zwei Erklärungen an: Zum einen enthalten die Angstkonzept-Definitionen trotz der angestrebten und wohl auch erzielten Beobachtungsnähe ein erhebliches Maß an Komplexität, denn jedes Konzept enthält ja zahlreiche Einzelaspekte (Hinweise auf ...), die von verschiedenen Beurteilern auch unterschiedlich gewichtet und unterschiedlich zusammengefaßt werden können. Zum anderen wurden komplette Behandlungsstunden zu je ca. 10 bis 15 maschinengeschriebenen Seiten als „scoring unit“ (Dollard und Auld, 1959) verwendet, die jeweils eine Vielzahl von Hinweisen auf unterschiedliche Konzepte enthalten. Diese Komplexität der Beurteilungseinheit mag von verschiedenen Beurteilern ebenfalls unterschiedlich gehandhabt worden sein. Zusammenfassend läßt sich mit Cohen (1965) feststellen, daß sowohl die Komplexität der Konzeptualisierungen wie auch die Komplexität des Beurteilungsmaterials zu mittleren Inter-Beurteiler-Übereinstimmungen führen.

Was die Vereinfachung der Konzeptualisierungen angeht, so könnte man daran denken, die Operationalisierungen in kleinste inhaltlich

<sup>9</sup> Eine Publikation der inzwischen fortgeführten Arbeiten befindet sich in Vorbereitung (persönliche Mitteilung von Meyer).

sinnvolle<sup>10</sup> Untereinheiten (Items) zu zerlegen und jedes Item für sich danach zu bewerten, ob es in der jeweiligen „scoring unit“ dynamisch wirksam ist oder nicht. Nach dieser Beurteilungsprozedur könnte rein rechnerisch das Gesamtkonzept in seiner Intensität durch Summation der einzelnen Items rekonstruiert werden. Dieses Vorgehen sollte zu höheren Inter-Beurteiler-Übereinstimmungswerten führen.

Wir verfolgten den zweiten aufgeführten Ansatz und versuchten, die Inter-Beurteiler-Übereinstimmung anzuheben, indem wir die Komplexität des Textmaterials reduzierten. Einmal wurde die Länge der „scoring unit“, bislang die Dialogsequenz einer kompletten Behandlungsstunde, radikal verkleinert. Um in diesem Rahmen die inhaltlich bedeutungsvollen Texteinheiten nicht übermäßig zu trennen, wählten wir als Beurteilungseinheit Äußerungen des Patienten, die nicht vom Analytiker unterbrochen wurden. Auf Patientenäußerungen beschränkten wir uns deshalb, weil anzunehmen ist, daß sich die Angstthematiken beim Patienten erlebnisnäher und weniger theoriebezogen ausdrücken als beim Psychoanalytiker. Zum anderen zielten wir auf möglichst prägnante und klare Beurteilungseinheiten im Sinne der beschriebenen Angstkonzepte ab.

### *Angsteinstufungen anhand von Segmenten*

Aus den Verbatimprotokollen der 500 Behandlungsstunden wurden ca. 200 Patientenäußerungen ausgesucht, welche thematisch die Kastrations-, Trennungs-, Schuld-, bzw. Beschämungsangst repräsentieren<sup>11</sup>. Diese Äußerungen wurden aus dem Gesamtkontext der Stunden herausgenommen und mit Hilfe der beschriebenen vierstufigen Intensitätsskala hinsichtlich der konzeptuellen Prägnanz eingestuft. Für jede der vier Angstformen wurden die 10 prägnantesten Äußerungen ausgewählt und als Einstufungsmaterial in den im folgenden beschriebenen Untersuchungen verwandt<sup>12</sup>. Für jede Angstform sei ein charakteristisches Beispiel angeführt:

<sup>10</sup> „Sinnvoll“ bezieht sich hier auf klinisch-theoretische Relevanz im Rahmen der Psychoanalyse.

<sup>11</sup> Die Untersuchung des Konzepts „Diffuse Angst“ ist fallengelassen worden, weil uns in dieser Teiluntersuchung in erster Linie die Beurteilung *objektbezogener* Angstformen interessierte.

<sup>12</sup> Die vollständige Liste der 40 Angstäußerungen kann bei den Autoren angefordert werden.

1014 *H. Thomä und andere*

Tabelle 4

*Operationale Definition*

*Textbeispiel*

**Trennungsangst**

Hinweise auf Verlassen, Verlassenheit, Einsamkeit, Verstoßung, Verlust von Liebe oder Liebesobjekt oder eine damit verbundene Drohung.

Ich weiß nicht, weshalb mir's schlecht wird, wenn ich mich ein paar Schritte von hier wegbewege.  
Ach, das ist natürlich eigentlich alles — das Nichthinausgehen. Es ist alles drin, was irgendwo meine Krankheit ausmacht. Das ist eigentlich zum Verzweifeln, weil, ich will ja hinaus.

**Kastrationsangst**

Hinweise auf Verletzung, physische Beschädigung oder Angst vor Verletzung, weiterhin Hinweise, die eine Bedrohung der Potenz und des Könnens überhaupt repräsentieren.

Ich hab ja die Zeit nicht genützt, vielleicht hätte ich mehr aus Ihnen machen können, aber es ist mir mal wieder schiefgelaufen, hab' wieder nichts geleistet. Es trifft zunächst einmal mich, denn ich will ja etwas können, denn ich hab' was nicht gekonnt, nämlich hinausgehen, arbeiten, Mädchen, ich bin's ja, der nichts kann. Mein Körper ist futsch.

**Schuldangst**

Hinweise auf Kritik, moralische Verurteilung, Verdammung, Verurteilung auf Schuld oder Bedrohung durch Schuld.

Ja, dem ist ein Mann ins Auto gelaufen und das hat mich dann wieder erschreckt, weil mir der Gedanke natürlich wieder unerträglich ist, mir könnte so etwas passieren, daß ich jemand da . . .

**Beschämungsangst**

Hinweise auf das Gefühl der Lächerlichkeit, der Kränkung, des Ungenügens, der Scham und Erniedrigung.

Was interessiert mich das. Ich werde jedenfalls von anderen Leuten schräg angesehen z. B., und das ertrag ich nicht. Ich kann mich ja auch dann nicht irgendwie verteidigen. An dem Beispiel zeigt sich ja auch alles, alle möglichen anderen Ängste. Ich habe Angst, daß ich mich wieder dumm anstelle. Ich habe Angst, daß am Ende zu wenig rausguckt.

Dem Kliniker mag sich hier das Problem stellen, inwieweit solche kurzen Textausschnitte ohne den klinischen Kontext eine angemessene Beurteilung zulassen. Hier kann man unterscheiden zwischen dem Kontext der Behandlungsstunde oder eines Abschnitts dieser Stunde und dem Kontext des Falles. Der Kontext des Falles mag von Bedeutung sein, um beurteilen zu können, welche konnotativen Bedeutungen ein Patient an bestimmte Personen, Ereignisse oder Begriffe knüpft. In dem

hier vorliegenden Fall aber kann der Kontext des Falles unberücksichtigt bleiben, da die Auswahl der einzelnen Äußerungen sich an inter-individuell geltenden Konzepten orientierte.

Anders ist es bei dem Kontext der Stunde. Auf die Frage, was der kritisierende Kliniker von ihm erwarten mag, läßt sich antworten, daß der Kontext weitere Hinweise auf die sich im Item bereits abzeichnende Angstform bringen und so das endgültige Urteil erleichtern und absichern könnte. Es bleibt jedoch fraglich, ob der Kontext diese Funktion erfüllt. Hierzu sei kurz ein Experiment von Oskamp (1965) referiert:

Oskamp gab einer Reihe von Beurteilern Material über einen Fall zu lesen, ließ sie dann einige Fragen über diesen Patienten beantworten und die Sicherheit angeben, mit der sie ihre Urteile gefällt hatten. Die Untersuchung wurde in drei Durchgängen durchgeführt; von Durchgang zu Durchgang war der Umfang der Unterlagen über den Patienten größer. Die Validität der Beurteilungen des Falles wurde dadurch kaum verbessert, wohl aber stieg die Sicherheit, mit der die Beurteiler ihr Urteil abgeben, rapide an. In einem gewissen Ausmaß ist subjektive Sicherheit des Therapeuten zweifellos als positiv anzusehen. Berücksichtigt man jedoch, daß auch falsche Urteile mit einem starken Gefühl der Sicherheit abgegeben wurden und daß die Selbstsicherheit oft so groß war, daß sie ein Hinterfragen der abgegebenen Urteile verhinderte, ist die Rolle des Kontextes zumindest zu problematisieren.

Eine aus dem Kreis von Mitgliedern des Fachbereichs Psychotherapie-Psychosomatik der Universität Ulm zusammengestellte Beurteilergruppe sollte darüber Aufschluß geben, wie verlässlich die Auswahl der Items in bezug auf die Angstdefinitionen war. Insgesamt gab es eine recht gute Übereinstimmung (größer als 70 %) mit der ersten Beurteilergruppe. Zwölf Äußerungen jedoch schnitten in den Zuordnungsergebnissen erheblich schlechter ab; sie wurden nach der oben skizzierten Auswahlprozedur durch geeignetere Äußerungen ersetzt. Somit konnten 28 Äußerungen beibehalten werden; 12 neue kamen hinzu.

Infolge der Ausgangsfrage nach Übereinstimmungsleistungen in klinischen Fallseminaren stellen Psychoanalytiker die Zielgruppe dar, an der weitere Untersuchungen zur Konsensproblematik durchgeführt werden sollen, denn es geht ja vor allem um die Abklärung fachinterner Verständigungsmöglichkeiten. Allerdings stellt sich hier sofort die Frage nach der Notwendigkeit einer fachlichen Kompetenz, um bei dem dargestellten Beurteilungsmaterial zu einem hinreichenden Konsens zu gelangen. Wenn nämlich Angstkonzepte und Beurteilungssegmente sorgfältig ausgesucht, d. h. beobachtungsnah gestaltet werden, kann man davon ausgehen, daß in den Angstformen fundamentale Erlebnisbereiche — wie Sich-Schämen, Sich-verlassen, Sich-schuldig-Fühlen oder Sich-körperlich-bedroht-Fühlen etc. — angesprochen sind. An diesen Erlebnissen haben alle Menschen mehr oder weniger gleichen Anteil,

verfügen in bezug auf sie über einen gemeinsamen Erfahrungshorizont. Psychoanalytische Experten sind darüber hinaus in der Lage, die in den sprachlichen Äußerungen enthaltenen latenten Bedeutungen zu erkennen, indem sie die zahlreichen, phänomenologisch beschreibbaren neurotischen Angstsituationen psychogenetisch auf eine kleine Zahl prototypischer Gefahrensituationen zurückführen.

Der zunächst unverständliche Anteil von neurotischen Ängsten besteht in der Unangemessenheit von tatsächlicher und vorgestellter Gefahrensituation und der daran entwickelten Angst. Durch psychogenetische Reduktion werden neurotische Ängste unter Hinzunahme eines theoretischen Rahmens erklärt und damit zugleich verständlich gemacht.

Dieser spezifisch psychoanalytische Beitrag hat sich besonders in der Formulierung der operationalisierten Angstdefinitionen, aber auch in der Auswahl der Textbeispiele niedergeschlagen. Es kann nämlich davon ausgegangen werden, daß bei der Auswahl sowohl die Vielfalt der phänomenalen Erscheinungsformen von Angst wie auch ihre psychogenetischen Entstehungsbedingungen als ein wesentliches Ordnungsprinzip berücksichtigt wurden. Nachdem eine solche Vorleistung erbracht ist, können Psychoanalytiker im Beurteilungsprozeß ihre fachliche Kompetenz nicht mehr wirksam einsetzen, d. h. der Informationsvorsprung, den Experten auf Grund ihres theoretischen Wissens gegenüber Laien haben, kann nicht angemessen zum Tragen kommen. Es ist im Gegenteil sogar zu vermuten, daß Psychoanalytiker auf Grund ihrer klinischen Erfahrung dazu neigen, die offenkundigen Angsthinweise in den Textbeispielen gering zu achten und die latenten Angstinhalte übermäßig stark gewichtet in die Beurteilung einfließen zu lassen. Bedenkt man den Umstand, daß unsere operationalen Angstdefinitionen gerade auf manifeste Inhalte abheben, so ist zu erwarten, daß in den Beurteilungsergebnissen psychoanalytische Experten jedenfalls nicht besser abschneiden werden als Laien.

Die Frage nach der Inter-Beurteiler-Übereinstimmung hinsichtlich der 40 Äußerungen wurde an zwei Personengruppen untersucht: Einmal nahmen 25 erfahrene Psychoanalytiker aus dem gesamten Bundesgebiet als Experten, zum anderen 45 Nicht-Analytiker aus Ulm als Laien an der Untersuchung teil. Die Gruppe der Laien setzt sich aus 15 medizinischen Dokumentationsassistentinnen, 15 Sekretärinnen des Fachbereichs Psychotherapie der Universität Ulm und aus 15 Medizinstudenten in den klinischen Semestern zusammen.

Die Versuchsteilnehmer wurden mit einer standardisierten Instruktion aufgefordert, für jede der 40 vorgelegten Äußerungen anzugeben,

welche der vier operational definierten Angstthematiken die jeweils vorherrschende sei. Die Beantwortung dieser forced-choice-Aufgabe erfolgte auf einem vorbereiteten Protokollbogen, der durch seine Gestaltung den Versuchsteilnehmern keineswegs nahelegte, jede der vier Angstformen gleich häufig zu wählen. Auch die Instruktion enthielt keinen Hinweis darauf, daß die vier Thematiken in den 40 Äußerungen gleich häufig repräsentiert waren.

Den Analytikern wie auch den Dokumentaristinnen und den Sekretärinnen wurden die Operationalisierungen mit den Benennungen der Angstformen (also z. B. Trennungsangst, Schuldangst usw.) vorgelegt; bei der Gruppe der Studenten wurde statt von den benannten Angstformen nur von Thematiken in neutraler Bezeichnung gesprochen (z. B. Thematik 1, Thematik 2, usw.). Die Studenten wußten also nicht, daß es sich bei dieser Untersuchung um Einstufungen von Angstformen handelt, und auch nicht, welche theoretischen Konzepte zugrunde lagen. Diese Variation der Versuchsdurchführung geht auf folgende Überlegung zurück:

Die Vorgabe der Kurzbezeichnung der untersuchten Konzepte, wie z. B. Kastrationsangst, aktualisiert vielleicht gerade bei psychoanalytischen Laien Einstellungshaltungen, welche individuellen oder auch allgemein menschlichen Erlebnisbereichen entspringen. Die hieraus entstehende Interferenz mit den beobachtungsnahen Konzeptdefinitionen könnte sich in einer verringerten Inter-Beurteiler-Übereinstimmung niederschlagen. Da dies infolge der geänderten Instruktion bei den Studenten nicht eintreten kann, ist bei dieser Gruppe ein höherer Konsens zu erwarten.

### *Ergebnisse*

Die folgende Tabelle, untergliedert nach den vier Angstthematiken und nach den Personengruppen, gibt Aufschluß über den Mittelwert ( $\bar{x}$ ) und die Standardabweichung ( $s$ ) der Inter-Beurteiler-Übereinstimmungen. Als Kriterium für die Richtigkeit der Zuordnungen gilt die Festlegung der thematischen Zugehörigkeit der 40 Patientenäußerungen, wie sie von der Autorengruppe vorgenommen worden war. Daneben ist ein weiteres Maß für die Konsensleistung innerhalb der Beurteilergruppe angegeben, das von der Festlegung der Autorengruppe unabhängig ist. Für die quantitative Bestimmung einer solchen Intra-Gruppen-Übereinstimmung ohne Berücksichtigung der „Richtigkeit“ der Beurteilungen



verwandten wir den Konkordanzkoeffizient  $W$  von Kendall (vgl. Siegel, 1956).

Tabelle 5

*Mittelwert ( $\bar{x}$ ), Standardabweichung ( $s$ ) und Konkordanzkoeffizient ( $w$ ) der Inter-Beurteiler-Übereinstimmungen*

	Trennungs- angst	Schuld- angst	Beschämungs- angst	Kastrations- angst	Gesamt
Gruppe der Analytiker	$\bar{x} = 8.4$ $s = 1.4$ $w = .62$	$\bar{x} = 7.4$ $s = 1.9$ $w = .59$	$\bar{x} = 6.8$ $s = 1.8$ $w = .52$	$\bar{x} = 5.9$ $s = 1.8$ $w = .62$	$\bar{x} = 7.1$ $s = 1.7$ $w = .59$
Gruppe der Dokumentations- assistentinnen	$\bar{x} = 7.4$ $s = 1.5$ $w = .61$	$\bar{x} = 6.9$ $s = 1.2$ $w = .68$	$\bar{x} = 6.7$ $s = 1.8$ $w = .73$	$\bar{x} = 5.2$ $s = 1.9$ $w = .54$	$\bar{x} = 6.6$ $s = 1.6$ $w = .64$
Gruppe der Sekretärinnen	$\bar{x} = 8.6$ $s = 1.1$ $w = .61$	$\bar{x} = 6.5$ $s = 1.9$ $w = .61$	$\bar{x} = 6.5$ $s = 1.7$ $w = .79$	$\bar{x} = 5.1$ $s = 2.2$ $w = .51$	$\bar{x} = 6.7$ $s = 1.8$ $w = .63$
Gruppe der Studenten	$\bar{x} = 6.8$ $s = 1.3$ $w = .67$	$\bar{x} = 6.8$ $s = 1.5$ $w = .57$	$\bar{x} = 6.2$ $s = 1.1$ $w = .73$	$\bar{x} = 4.3$ $s = 1.8$ $w = .62$	$\bar{x} = 6.0$ $s = 1.4$ $w = .65$
Gesamt	$\bar{x} = 7.7$ $s = 1.3$ $w = .63$	$\bar{x} = 6.9$ $s = 1.6$ $w = .61$	$\bar{x} = 6.6$ $s = 1.6$ $w = .69$	$\bar{x} = 5.1$ $s = 1.9$ $w = .57$	$\bar{x} = 6.6$ $s = 1.6$ $w = .63$

Die in den mit  $\bar{x}$  bezeichneten Zeilen aufgeführten Werte geben an, wie viele von den je 10 Äußerungen durchschnittlich „richtig“ der jeweiligen Thematik zugeordnet wurden. Da die je 10 Äußerungen sich unter der statistischen Nullhypothese gleichmäßig auf die vier Thematiken verteilen sollten, wird unter dieser Annahme eine durchschnittliche richtige Zuordnungshäufigkeit von 2.5 erwartet. Die in der Tabelle aufgeführten Werte weichen von diesem Erwartungswert statistisch gesichert ab ( $P < .01$ ), wenn man von den Mittelwerten für die Kastrationsthematik der Dokumentaristinnen und der Sekretärinnen ( $P < .05$ ) und von dem Mittelwert für die Kastrationsthematik der Studenten (nicht signifikant) absieht. Alle vier Versuchsteilnehmergruppen weisen eine Gemeinsamkeit darin auf, daß stets die Höhe der „richtigen“ Zuordnungen von der Trennungsangst über die Schuldangst und Beschämungsangst zur Kastrationsangst abnimmt.

Die aufgeführten Konkordanzkoeffizienten sind statistisch gesichert

( $P < .01$ ), d. h. alle vier Personengruppen kommen für jede Angstform in der Beurteilung der Patientenäußerungen zu einer eindrucksvollen Einigung. Obwohl Kendalls Koeffizient die folgende Aussage nicht erlaubt, ist an der Verteilung der einzelnen Zuordnungen zu erkennen, daß diese Einigung ausschließlich auf die „richtigen“ Zuordnungen der Äußerungen zurückgeht. Wertet man die Intragruppenübereinstimmung bezüglich der „falsch“ zugeordneten Äußerungen mit dem gleichen Verfahren aus, so zeigt sich, daß die Gruppen in ihren Fehlzuweisungen nun uneinheitlich verfahren. Mit anderen Worten: Bezüglich der „Richtig-Zuweisungen“ läßt sich in allen Gruppen ein auffallend hoher Konsens herstellen, die Zuweisungen der restlichen, „falsch“ zugeordneten Äußerungen hingegen werden unsystematisch und heterogen auf die jeweils restlichen drei Angstkategorien verteilt.

Als erstes bedeutendes Ergebnis kann also festgehalten werden, daß bei einer geeigneten beobachtungsnahen Operationalisierung theoretisch abgeleiteter Angstkonzepte und bei entsprechender sorgfältiger Auswahl von Textbeispielen ein beeindruckend hoher Konsens zwischen Klinikern wie auch zwischen Laien erzielt werden kann. Dieser Befund ermutigt zu weiteren Forschungen, die schließlich in eine empirisch fundierte Validierung dieser Konstrukte einmünden können.

Weiterhin stellt sich die Frage, ob die vier thematischen Gruppen unterschiedlich übereinstimmend beurteilt worden sind, und ob sich psychoanalytische Experten in der Zahl der „richtigen“ Zuweisungen von den Laien abheben. Die bereits weiter oben erwähnte Überlegung, wonach die Studenten auf Grund der geänderten Instruktion zu höheren Inter-Beurteiler-Übereinstimmungen als die anderen Laiengruppen gelangen sollten, soll ebenfalls quantitativ überprüft werden. Die Formulierung dieser Fragestellungen weist auf einen zweifaktoriellen varianzanalytischen Ansatz hin, dessen Resultate im folgenden erläutert werden sollen.

Tabelle 6

*Zweifaktorielle Varianzanalyse*

Quelle der Variation	Freiheitsgrade	Quadratsummen	Mittlere Quadratsummen	F	P
Pers.-Gruppen	3	2376.8	792.2	1.87	
Angstformen	3	14965.8	4988.6	11.80	.01
Wechselwirkung	9	1532.4	170.2	0.40	
Innerhalb	144	60872.7	422.7		
Gesamt	159	79747.8			

Wie aus dieser Tabelle hervorgeht, sind nur die Unterschiede zwischen den vier Angstthematiken statistisch gesichert; keine bedeutsamen Differenzen bestehen hinsichtlich der Personengruppen; auch eine Wechselwirkung zwischen Personengruppen und Angstformen ist nicht nachweisbar.

Wie der Duncan-Test (vgl. Mittenecker, 1965) ausweist, ist die Übereinstimmungsleistung hinsichtlich der Trennungsangstthematik am größten, bezüglich der Kastrationsangst am geringsten. Dieser Sachverhalt kann ebenfalls der Tabelle 5 entnommen werden. Zur weiteren Veranschaulichung haben wir alle 40 Textbeispiele nach der Zahl der erreichten „richtigen“ Zuordnungspunkte über alle Versuchsteilnehmer hinweg in eine gemeinsame Reihenfolge gebracht.

Die folgende Aufstellung enthält das Resultat dieser Rangbildung, wobei ein niedriger Rangwert einer hohen Übereinstimmung entspricht und umgekehrt.

Kastrationsangst	Rang	8, 20, 25, 26, 29, 30, 33, 38, 39, 40
Beschämungsangst	Rang	4, 7, 11, 12, 22, 23, 31, 34, 36, 37
Schuldangst	Rang	3, 6, 10, 16, 18, 21, 24, 27, 32, 35
Trennungsangst	Rang	1, 2, 5, 9, 13, 14, 15, 17, 19, 28

Wie läßt sich nun verstehen, daß die Textbeispiele für Kastrationsangst seltener „richtig“ beurteilt wurden als die Trennungsangstäußerungen? Theoretisch läßt sich zunächst sagen, daß die Kastrationsangst neben der Triebkomponente noch eine Über-Ich-Komponente und eine phallich-narzißtische Komponente hat, die wir — wie rückblickend festzuhalten ist — bei der Konzeptualisierung dieser Angstform nicht genügend in Rechnung gestellt haben. Die Beurteiler nehmen bei den Textbeispielen häufig auch Aspekte von Schuld und Beschämung wahr, was durch die ödipale Konfliktkonfiguration, die ja diesen drei Themenbereichen gemeinsam ist, sowohl auf der Ebene der Konzeptbildung wie auch auf der Ebene der operationalen Definitionen deutlich zum Ausdruck kommt. Anders dagegen bei Trennungsangst, die durch ihr prä-ödipales Konfliktmuster davon deutlich abgrenzbare Inhalte zeigt.

Eine eher methodenkritische Alternative für die Interpretation des Varianzanalytischen Befundes mag dahin gehen, daß die Textbeispiele für Trennungsangst prägnanter ausgefallen sind als diejenigen für Kastrationsangst. Allerdings stellt sich hier sofort die Frage, warum dieser Analyse-Patient nur so unklare Kastrationsangstbeispiele und so prägnante Trennungsangstbeispiele produziert hat; sieht man sich etwa die

körperliche Symptomatik an (Tachykardie, Herzängste etc.), so wird das jedem Kliniker einsichtig sein.

Wie die varianzanalytische Überprüfung ergeben hat, gelangen psychoanalytische Experten nicht zu höheren Übereinstimmungsergebnissen als psychoanalytische Laien. Bei der Berücksichtigung unserer Vorüberlegungen ist dieses Ergebnis allerdings nicht überraschend. Dieser empirische Befund erhärtet vielmehr die Überlegung, daß in die Beurteilerleistung keine spezifisch psychoanalytische Kompetenz eingegangen ist, welche offenbar bereits in der Art der Konzeptualisierung, der Operationalisierung sowie in der Materialauswahl enthalten ist.

Die weiter oben erwähnte Hypothese, wonach die Studentengruppe auf Grund der geänderten Instruktion zu höheren Übereinstimmungsergebnissen als die anderen Laiengruppen gelangen sollte, ließ sich nicht bestätigen. Die vermutete Diskrepanz zwischen den beobachtungsnahen Konzeptionsdefinitionen und den individuellen oder auch allgemein menschlichen Erlebnissen hat sich zumindest in dieser Einstufungsuntersuchung nicht gravierend ausgewirkt. Vielmehr scheinen sich die Versuchsteilnehmer bei ihrer Beurteilung ziemlich eng an die vorgegebenen operationalen Hinweise auf die Konzepte gehalten zu haben.

Das für Psychoanalytiker vielleicht überraschende Resultat, bei dieser Aufgabenstellung nicht erfolgreicher abgeschnitten zu haben als die Laien, soll noch näher untersucht werden. Von der Überlegung ausgehend, daß Textbeispiele mit einem höheren Theoriegehalt von der Expertengruppe eigentlich zutreffender beurteilt werden sollten als von den Laiengruppen, schätzten wir jedes einzelne Textbeispiel nach dem Ausmaß ein, in dem Teile der psychoanalytischen Angsttheorien in ihm enthalten sind. Es wäre nun zu vermuten, daß bei denjenigen Textbeispielen mit einem hohen Theorieanteil bzw. mit einem hohen Inferenzgrad die Experten den Laien gegenüber im Vorteil sind und somit zu besseren Einstufungsergebnissen gelangen; bei den Textbeispielen mit geringem Inferenzgrad sollten demzufolge keine erheblichen Unterschiede zwischen Experten und Laien nachweisbar sein. Eine varianzanalytische Überprüfung konnte diese Vermutungen allerdings nicht bestätigen.

### *Klinische Beschreibungen extremer Äußerungen*

Für eine klinische Beschreibung einiger extrem beurteilter Äußerungen haben wir nach folgenden Gesichtspunkten eine Auswahl getroffen:

1. Die Äußerung, welche am besten über alle Gruppen zugeordnet wurde;
2. zwei Äußerungen, welche eine geringe Übereinstimmung erzielten;
3. eine Äußerung, die von Analytikern gut, von den psychoanalytischen Laien eher schlecht zugeordnet wurde, und schließlich
4. eine Äußerung, die von den psychoanalytischen Laien erheblich zutreffender als von der Analytikergruppe eingestuft wurde.

1022 *H. Thomä und andere*

1. Äußerung:

„Ja, und dann, wenn's mir immer so schlecht ist wie heute mittag, da hab' ich immer den Wunsch, nach irgend jemand zu schreien. Ich muß mich da immer zusammennehmen, daß ich nicht schreie; irgendwie nicht alleingelassen werden wollen oder so was fällt mir noch dazu ein.“

Diese Äußerung hat über alle Gruppen eine hundertprozentige Übereinstimmung erzielt. In dieser Äußerung verbinden sich szenische wie auch Textelemente zu einer Aussageprägnanz, die ganz manifeste Trennungangsthinweise enthält und kaum eine Alternativzuordnung zuläßt. Angst vor der Trennung drückt sich in dem Wunsch, „nach irgend jemand zu schreien“, aus wie in der Angst, „nicht alleingelassen zu werden“.

2. a) Äußerung:

„Ich habe die Überzeugung, daß das, was ich mache, letzten Endes Scheiße wird. Ich bin bei allem abhängig von dem, was andere Leute eben davon denken. Ich habe Angst, daß jemand kritisiert, was ich mache.“

Diese Äußerung hat wohl deshalb über alle Gruppen eine eher geringe Übereinstimmung gebracht, weil in ihr Textelemente enthalten sind, die sowohl den Beschämungs- als auch den Schuldaspekt ansprechen. Für viele, die sie der Schuldangstkategorie zugewiesen haben, war sicher das Wort „Kritik“ der entscheidende Hinweis. Im übertragenen Sinne ist aber auch von dem Gefühl der Beschämung und Erniedrigung die Rede. Zum Beispiel: „... daß das, was ich mache, letzten Endes Scheiße wird.“ Interessant ist, daß die Fachgruppe diese Äußerung unter den Schuldangstbeispielen an erste Stelle rückte; sie war also der Meinung, daß diese Äußerung sehr einfach zuzuordnen sei. Möglicherweise ließ sich die Fachgruppe mehr von dem indirekten Textgehalt als von Definitionsaspekten der Konzepte leiten.

2. b) Äußerung:

„... hatte ich für meinen Teil immer eine — sagen wir — körperfeindliche Haltung eingenommen, die wohl der Tatsache entsprang, daß ich, meiner sonstigen Haltung entsprechend, besonders in bezug auf sexuelle Dinge, eben den Körper als etwas Niederes empfinden mußte, gewissermaßen eine Abneigung züchten mußte.“

Diese Äußerung konnten insgesamt nur sechs Gruppenmitglieder mit einer Kastrationsthematik verknüpfen. Mehrheitlich schlugen die Beurteiler über alle Gruppen hinweg diese Äußerung der Beschämungsthematik zu (allein 17 Analytiker); vereinzelt wurden auch Schuldinhalte erkannt. In dieser Äußerung laufen in der Tat drei Schlüsseläußerungen nebeneinander her:

- Der Kastrationsaspekt: „Körperfeindliche Haltung eingenommen ... besonders in bezug auf sexuelle Dinge“
- Der Beschämungsaspekt: „Den Körper als etwas Niederes“
- Der Schuldaspekt: „Abneigung züchten mußte“ (womit wohl auf Selbstbestrafungstendenzen verwiesen wird).

### 3. Äußerung:

„Hinterher war's mir auch ab und zu übel. Aber ich kann nicht sagen, ob das eine Art Ekel war oder irgendwie aus einer Reue oder aus einem Schuldgefühl entspringt. Das kann ich nicht genau sagen. Ich hab auch seltsamerweise Selbstbefriedigung und normalen Koitus etwa gleich bewertet, mit gleichen Unwertgedanken belegt.“

Die Gruppe der Analytiker hat diese Äußerung in Übereinstimmung mit der Autorengruppe mehrheitlich der Schuldangst zugewiesen. Für die Gruppe der Laien muß wohl maßgeblich der sexuelle Aspekt (Selbstbefriedigung, Koitus) bei der erfolgten Zuweisung zur Kastrationsangst im Vordergrund gestanden haben.

### 4. Äußerung:

„Na, ich weiß nicht. Ich kann, weiß der Kuckuck, nichts damit anfangen. Das versteh' ich, was Sie sagen. Aber mit dem Verständnis wird's ja wohl nicht getan sein. Sicher weiß ich, daß das — das hatten wir ja früher schon besprochen —, daß es schlimm oder wenigstens für mich schlimm ist, mit dem Gefühl allein zu sein. Ich weiß auch, daß es schlimm ist, wenn ich sehe, daß das nicht erwidert wird, oder daß ich die Zuneigung von jemand anderem verloren habe, oder daß ich immer den Wunsch habe, meine Empfindungen zu drosseln. Aber ich weiß ja jetzt, daß es unklug ist, so zu verfahren. Aber ich meine, der Schmerz wird immer bleiben. Ich glaube nicht, daß man das beiseite schaffen kann. Es geht einfach nicht. Sie haben ja bemerkt, ich glaube, daß, wenn ich mich recht entsinne, haben Sie das sogar hier beispielsweise konstatiert, daß ich versucht habe, der Empfindung, daß ich Sie liebhave, mehr Raum zu geben, beispielsweise durch die Wahl der Worte, daß es mir aber effektiv nicht gelungen ist, die Intensität zu steigern; obwohl ich einerseits die gefühlsmäßige Möglichkeit sehe und auch, sagen wir, die Sehnsucht habe, nun jemand liebhaben zu können, daß es aber trotzdem nicht geht, abgeschnitten. Aber ich habe eben den Eindruck, nun, ich weiß ja nicht, entweder hab' ich Sie bis jetzt mißverstanden, oder ich weiß nicht, was ich damit anfangen soll, oder eben die dritte Möglichkeit, da ist irgend etwas anderes, von dem ich nicht weiß, was es mir so schlimm macht, jemanden gern zu haben.“

Diese Äußerung ist von der Gruppe der Dokumentationsassistentinnen und Sekretärinnen eindeutig der Trennungsangst zugeordnet worden, während nur 14 von 24 Analytikern Trennungsangst-Zuordnungen getroffen haben. Es ist zugleich die längste der 40 Patientenäußerungen; für den psychoanalytisch geschulten Beobachter sind sicherlich implizit und explizit zahlreiche psychodynamische Themen angesprochen. Das Thema des Verlassenseins und Alleinseins interferiert mit dem Thema von Nehmen und Geben; der Wunsch, jemand gerne zu haben, interfe-

riert mit der Furcht davor. Der Experte wird sich hier schwer getan haben, welches Thema er bevorzugt gewichten will, während die Laien bei der verwirrenden Vielfalt von Themen wohl den manifesten Hinweis auf das Verlassensein als deutlichsten Hinweis für eine der vorgegebenen Konzeptdefinitionen angesehen haben.

### *Abschließende Diskussion*

In den bisher erwähnten empirischen Untersuchungen wurden stets die Angstkonzepte als Konstrukte vorgegeben. Will man nun zu einer *Validierung* dieser Konstrukte gelangen, so versteht sich von selbst, daß auf die Vorgabe der Konzeptualisierungen verzichtet werden muß. Statt dessen lassen wir die Versuchspersonen die Einzeläußerungen in einer weiteren Untersuchung von Plewa et al. (1976) im sog. Tripelvergleich hinsichtlich der thematischen *Ähnlichkeit* beurteilen, ohne daß auf besondere Thematiken oder gar Angstformen hingewiesen wird. Sollte sich ein Äußerungs-Cluster herausbilden, das sich überwiegend z. B. aus Trennungsangst-Äußerungen zusammensetzt, so wäre hiermit ein wesentlicher Beitrag zur Konstruktvalidierung geleistet. Das Auffinden solcher Cluster geschieht mit entsprechenden mehrdimensionalen Raumanalysemethoden (wie z. B. Clusteranalyse und mehrdimensionalen Skalierungstechniken).

(Anschrift der Verff.: Prof. Dr. Helmut Thomä; Dr. phil. Hans-Joachim Grünzig, Dipl.-Psych.; Dipl.-Psych. Hermann Böckenförde und Priv.-Doz. Dr. med. Horst Kächele, Am Hochsträß 8 (Kuhberg), 7900 Ulm)

### *Summary*

Thomä et al.: *The consensus problem in psychoanalysis*. — After a thorough discussion of methodological questions concerning the problem of consensus in psychoanalysis (Part I) with special consideration of the arguments of Ricoeur and Lorenzer, and following a critical survey of the literature, two relevant empirical investigations are reported. In the first of these, the protocols of 55 sessions of a psychoanalytic treatment case were rated by means of rating scales constructed as operationalizations of anxiety concepts (shame anxiety, castration anxiety, guilt anxiety, separation anxiety). Three raters achieved average inter-rater agreement in line with values reported in the literature. In the second investigation, the session protocols were divided into separate patient-utterances. After a careful selection procedure, forty such utterances crystallized. These were presented to 25 psychoanalysts as well as to 45 laymen for evaluation with reference to concepts of anxiety. The fact

that experts and laymen attained high inter- and intra-group agreement suggests that the operationalization of these concepts in the evaluation procedure makes preconceptions related to theory dispensable. It may be noted that the selection of text units was confirmed by various groups; this may be viewed as an essential condition for the successful construct validation of concepts of anxiety.

# BIBLIOGRAPHIE

- Angel, K. (1972): The Role of the Internal Object and External Object in Object Relationships, Separation Anxiety, Object Constancy and Symbiosis. *Int. J. Psychoanal.* 53, 541—546.
- Apel, K. O. (1976 a): Transformation der Philosophie, Bd. I, Sprachanalytik, Semiotik, Hermeneutik. Frankfurt (Suhrkamp).
- (1976 b): Transformation der Philosophie, Bd. II, Das Apriori der Kommunikationsgemeinschaft. Frankfurt (Suhrkamp).
- Argelander, R., H. Bechtler, A. Durner und H. Thomä (1976): Ein kasuistischer Beitrag zur Erfolgsbeurteilung bei Gruppenpsychotherapien. *Psychotherapie und Med. Psychol.*, im Druck.
- Auerbach, A. H. und L. Luborsky (1968): Accuracy of Judgements of Psychotherapy and the Nature of the „Good Hour“. In: Shlien, J. M. (Hg.): *Research in Psychotherapy*. Am. Psychol. Ass. Washington D. C., Vol. III, 155—168.
- Bellak, L. und M. B. Smith (1956): An Experimental Exploration of the Psychoanalytic Process. *Psychoanal. Quart.* 25, 385—494.
- Bernfeld, S. (1932): Der Begriff der „Deutung“ in der Psychoanalyse. *Z. angew. Psychol.*, 42, 448—497.
- Cohen, R. (1965): Empirische Untersuchungen zur klinischen Urteilsbildung aufgrund psychologischer Tests. In: F. P. Hardesty und K. Eyferth (Hg.): *Forderungen an die Psychologie*, 139—152. Bern (Huber).
- Dollard, J. und F. Auld (1959): *Scoring Human Motives; a Manual*. New Haven (Yale Univ. Press).
- Eissler, K. R. (1958): Remarks on some Variations in Psychoanalytical Technique. *Int. J. Psychoanal.*, 39, 222—229.
- Freud, S. (1914): Zur Einführung des Narzißmus. GW X.
- (1916/17): Vorlesungen zur Einführung in die Psychoanalyse. GW XI.
- (1926 a): Hemmung, Symptom und Angst. GW XIV.
- (1926 b): Die Frage der Laienanalyse. GW XIV.
- (1938): Abriß der Psychoanalyse. GW XVI.
- Friedman, L. (1976): Problems of an Action Theory of the Mind. *Int. Rev. Psychoanal.* 3, 129—138.
- Garduk, E. L. und E. A. Haggard (1972): Immediate Effects on Patients of Psychoanalytic Interpretations. *Psychological Issues* 28, New York (Int. Univ. Press).
- Glover, E. (1952): Research Methods in Psychoanalysis. *Int. J. Psychoanal.*, 33, 403—409.
- Göttner, H. (1973): *Logik der Interpretation*. München (W. Fink).
- Goldberg, L. R. (1968): Simple Models or Simple Processes? Some Research on Clinical Judgments. *Amer. Psychol.*, 23, 483—496.
- Gottschalk, L. A. und G. C. Gleser (1969): The Measurement of Psychological States through the Content-Analysis of Verbal Behavior. Berkeley (Univ. Cal. Press).
- Haggard, E. A. (1958): *Intraclass Correlation and the Analysis of Variance*. New York (The Drydes Press).



1026 *H. Thomä und andere*

- Hamburg, D. A., M. A. Sabshin, F. A. Board, R. R. Grinker, S. J. Korchin, H. Basowitz, H. Heath und H. Persky (1958): Classification and Rating of Emotional Experiences. *Arch. Neurol. Psychiat.*, 74, 415—426.
- Hanly, Ch. und M. Lazerowitz (Hg.) (1970): *Psychoanalysis and Philosophy*. New York (Int. Univ. Press).
- Hoffmann, P. J., P. Slovic und L. G. Rorer (1968): An Analysis of Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment. *Psychol. Bull.*, 69, 338—349.
- Holt, R. R. (1958): Clinical and Statistical Prediction. A Reformulation and Some New Data. *J. Abnorm. Soc. Psychol.*, 56, 1—12.
- Kächele, H., H. Thomä und H.-J. Grünzig (1976): *Psychoanalytische Verlaufsforschung. Probleme, Methoden und Ergebnisse* (in Vorb.).
- Klein, G. S. (1976): *Psychoanalytic Theory. An Exploration of Essentials*. New York (Int. Univ. Press).
- Kline, P. (1972): *Fact and Fantasy in Psychoanalysis*. London (Methuen).
- Kris, E. (1956): On some Vicissitudes of Insight in Psychoanalysis. *Int. J. Psychoanal.*, 37, 445—455.
- Kuhn, Th. S. (1967): *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt (Suhrkamp).
- Kunz, H. (1975): Die Erweiterung des Menschenbildes in der Psychoanalyse Sigmund Freuds. In: H.-G. Gadamer und P. Vogler (Hg.): *Neue Anthropologie*, 6, 44—113.
- Laplanche, J. und Pontalis, J.-B. (1967): *Das Vokabular der Psychoanalyse*. Frankfurt (Suhrkamp) 1972.
- Lorenzer, A. (1970): *Sprachzerstörung und Rekonstruktion*. Frankfurt (Suhrkamp).
- (1974): *Die Wahrheit der psychoanalytischen Erkenntnis*. Frankfurt (Suhrkamp).
- Luborsky, L., H. Graff, S. Pulver und H. Curtis (1973): A Clinical Quantitative Examination of Consensus on the Concept of Transference. *Arch. Gen. Psychiat.*, 29, 69—75.
- Mahler, M. S. (1958): On Two Crucial Phases of Integration of the Sense of Identity: Separation — Individuation and Sexual Identity. *J. Am. Psychoanal. Ass.*, 6, 131—142.
- Meehl, P. E. (1954): *Clinical vs. Statistical Prediction*. Minneapolis (Minnesota Univ. Press).
- Meyer, A. E., H. Otte, H.-J. Seeberger, H. Speidel und R. Zenker (1969): Die Inter-Beobachter-Übereinstimmung für die psychoanalytische Einordnung von Charakter- und Verhaltensbeschreibungen. *Psyche*, 11, 824—837.
- Mintz, J. und L. Luborsky (1971): Segments Versus Whole Sessions: Which is the Better Unit for Psychotherapy Research? *J. Abnorm. Psychol.* 78, 180—191.
- Mittenecker, E. (1965): *Planung und statistische Auswertung von Experimenten*. 9. Aufl. Wien (Deuticke).
- Modell, A. H. (1975): The Ego and the Id: 50 Years Later. *Int. J. Psychoanal.*, 56, 57—68.
- Nagera, H. (1969/70): *Psychoanalytische Grundbegriffe. Eine Einführung in Sigmund Freuds Terminologie und Theorienbildung*. Frankfurt (Fischer) 1974.
- Oskamp, S. (1965): Overconfidence in Case-Study Judgements. *J. Consult. Psychol.*, 29, 261—265.
- Piers, G. und M. B. Singer (1953): *Shame and Guilt. A Psychoanalytic and a Cultural Study*. Springfield, Ill. (C. C. Thomas).
- Plewa, A., H.-J. Grünzig und H. Kächele (1976): *Zur Dimensionalität von Angstäußerungen aus einer psychoanalytischen Behandlung* (in Vorb.).
- Popper, K. (1958): *Die offene Gesellschaft und ihre Feinde. Band II: Falsche Propheten*. Bern (Franke).

- Ramzy, I. (1974): How the Mind of the Psychoanalyst Works: An Essay on Psychoanalytic Inference. *Int. J. Psychoanal.*, 55, 543—550.
- und H. Shevrin (1976): The Nature of the Inference Process in Psychoanalytic Interpretation: A Critical Review of the Literature. *Int. J. Psychoanal.*, 57, 151—159.
- Rapaport, D. (1960): *Die Struktur der psychoanalytischen Theorie*. Stuttgart (Klett) 1966.
- Ricoeur, P. (1965): *Die Interpretation*, Frankfurt (Suhrkamp) 1969.
- (1969): *Hermeneutik und Psychoanalyse*, München (Kösel) 1974.
- Sandler, J. und W. G. Joffe (1967): Die Persistenz in der psychischen Funktion und Entwicklung, mit besonderem Bezug auf die Prozesse der Fixierung und Regression. *Psyche*, 21, 138—151.
- , C. Dare und A. Holder (1973): *Die Grundbegriffe der psychoanalytischen Therapie*. Stuttgart (Klett).
- Schafer, R. (1973): Action: Its Place in Psychoanalytic Interpretation and Theory. *Ann. Psychoanal.*, 1, 159—196, New York (Quadrangle/The New York Times Book Co.).
- Schlesinger, H. J. (1974): Problems of Doing Research on the Therapeutic Process. *J. Am. Psychol. Ass.*, 22, 3—13.
- Seitz, P. F. D. (1966): The Consensus Problem in Psychoanalytic Research. In: Gottschalk, L. A. und A. H. Auerbach (Hg.): *Methods of Research in Psychotherapy*. New York (Appleton-Century-Crofts), 209—225.
- Shakow, D. (1960): The Recorded Psychoanalytic Interview as an Objective Approach to Research in Psychoanalysis. *Psychoanal. Quart.*, 29, 82—97.
- Siegel, S. (1956): *Non-Parametric Statistics for the Behavioral Sciences*. New York (McGraw-Hill).
- Strupp, H. H., J. B. Chassan und J. A. Ewing (1966): Toward the Longitudinal Study of the Psychotherapeutic Process. In: L. A. Gottschalk und A. H. Auerbach (Hg.): *Methods of Research in Psychotherapy*. New York (Appleton-Century-Crofts).
- und J. V. Williams (1973): Some Determinants of Clinical Evaluations. In: H. H. Strupp: *Psychotherapy: Clinical, Research, and Theoretical Issues*. New York (Jason Aronson, Inc.).
- Taylor, Ch. (1975): *Erklärung und Interpretation in den Wissenschaften vom Menschen*. Frankfurt (Suhrkamp).
- Thomä, H. (1976): Über die Identität und das Selbstverständnis des Psychoanalytikers. *Psyche* (im Druck).
- und H. Kächele (1973): Wissenschaftstheoretische und methodologische Probleme der klinisch-psychoanalytischen Forschung. *Psyche*, 27, 205—355.
- Waelder, R. (1926): Über schizophrenes und schöpferisches Denken. *Int. Z. Psychoanal.*, 12, 298—308.
- (1970): Observation, Historical Reconstruction and Experiment: An Epistemological Study. In: Hanly und Lazerowitz, 280—326.
- Wallerstein, R. und H. Sampson (1971): Issues in Research in the Psychoanalytic Process. *Int. J. Psychoanal.*, 52, 11—50.