

Bericht aus dem Psychologischen Institut der Universität

Heidelberg

Gerhard SCHNEIDER

REFLEXIVITÄT ALS GRENZPROBLEM EINER KOGNITIVEN PSYCHOLOGIE

August 1983

(Erscheint in L. Eckensberger und E. Lantermann (Hrsg.):
"Emotion und Reflexivität". Göttingen: Hogrefe).

DISKUSSIONSPAPIER Nr. 37

GERHARD SCHNEIDER

REFLEXIVITÄT ALS GRENZPROBLEM EINER KOGNITIVEN PSYCHOLOGIE

Psychologisches Institut
der Universität
Heidelberg

Juli 1983

1. Einleitung und Überblick

Die Entwicklung und der Aufschwung der modernen kognitiven Psychologie seit Ende der 50er Jahre sind unauflöslich verknüpft mit der Bezugnahme auf den Computer als programmgesteuertes informationsverarbeitendes System. Auf Grund der darin angelegten theoretischen wie methodisch-methodologischen Möglichkeiten wird zum einen gegen das Verdikt des strikten Behaviorismus die Beschäftigung mit komplexen 'inneren', 'mental' Vorgängen 'rehabilitiert'. Zum anderen bieten sich aber völlig neue expansive Möglichkeiten an: Löst man sich nämlich von der Vorstellung, daß der Computer im wesentlichen ein (Schnell-)Rechner zur Verarbeitung großer Zahlenmengen sei, so wird seine allgemeine Fähigkeit zur Symbolmanipulation erkennbar, er ist ein "physisch realisiertes symbolmanipulatives System" ("physical symbol system", Newell 1981, p. 38 ff.). Daraus erhellt das paradigmatische Potential dieser Maschinen (und zwar qua Programm!), sind doch zweckgeleitete menschliche Aktivitäten wie Denken, Planen, Problemlösen usw. wesentlich symbolvermittelt.

Es stellt sich natürlich die Frage nach der 'Rekonstruktions-Reichweite' dieses Paradigmas: Wie weit lassen sich psychologisch relevante menschliche Phänomene oder sogar der Mensch überhaupt mit Hilfe der Computer-Metapher (vgl.a. "computational metaphor", Böden 1981b) explizieren? "No limits"-Antworten sind im Prinzip beinhaltet, wenn bspw. im Rahmen der aus der kognitiven Psychologie, Artificial-Intelligence Forschung und anderen verwandten Richtungen entstehenden Metadisziplin "Cognitive Science" (vgl. Norman 1981) als Hypothese formuliert wird, daß die Klasse der rationalitätsfähigen Entitäten genau die der physischen symbolmanipulativen Systeme sei (Newell 1981, p. 72), oder wenn 'verstehen' mit 'in der Form eines Programms formulieren' gleichgesetzt wird, wie es nach G.A. Miller in der Psychologie zunehmend geschieht (Weizenbaum 1982, p. 211 f.). Dem stehen eine Reihe limitativer Versuche gegenüber, bspw. wenn gezeigt werden soll, daß der menschliche Geist prinzipiell nicht durch das Modell des Computers abbildbar sei (Lucas 1961), oder daß es bestimmte programmresistente Qualitäten - etwa Ambiguitätstoleranz - gebe (so bspw. Dreyfus 1972). Im deutschen Sprachraum ist G. Frey zu nennen, nach dem Reflexivität nicht in einem formalen axiomatischen System abbildbar ist. Demzufolge könne Maschinen kein Bewußtsein (charakterisiert durch Intentionalität und Reflexivität) zugesprochen werden (vgl. bspw. Frey 1965, 1975, 1980).

Die hier vorzubringenden Überlegungen zum Thema 'Reflexivität' gehören zur Klasse der letztgenannten limitativen Versuche, wobei eine enge Bezugnahme auf den Rahmen der kognitiven Psychologie angestrebt ist. Demzufolge geht es um die argumentative Begründung der folgenden These: Innerhalb des durch 'Reflexivität' angesprochenen Phänomenbereichs las-

sen sich Strukturen aufzeigen, die sich der Explikation im Rahmen der Computer-Metapher entziehen.

Nach einer Skizzierung des Problemfelds selbst (Abschn. 2) wird zunächst expliziert, was unter dem Konzept 'kognitive Psychologie' verstanden werden soll (Abschn. 3). Der dabei zentrale Begriff des Algorithmus erlaubt vor dem Hintergrund von Resultaten der mathematischen Logik und Metamathematik i.S. der obigen These die Schlußfolgerung, daß sowohl gewisse 'reflexiv durchsetzte' Einzelprozesse ('Beweisen' in bezug auf komplexere Problembereiche (Abschn. 4.1) wie auch 'Reflexivität überhaupt' (Abschn. 4.2) nicht als algorithmisch rekonstruierbar angesehen werden können. In einem dritten Schritt soll schließlich auf Grund sprachanalytischer Überlegungen (Schmitz, Castañeda) dargelegt werden, daß die Ich-Perspektive und damit Selbstbewußtsein in Programmen der angegebenen Art nicht im Vollsinn dieser Worterealisiert ist (Abschn. 4.3). Referenzpunkt der Überlegungen in Abschn. 4 ist das eine Maschine voll spezifizierende Programm, so daß prinzipiell von der Realisierung in einer Hardware abgesehen werden kann. Verläßt man diesen Bereich der kognitiven Psychologie i.e.S. und betrachtet hardware-realisierte lernende Systeme, so scheinen die vorstehenden Argumente nicht einfach übertragbar (Boden, Dennett). Deshalb wird versucht, einige andersgeartete Unwahrscheinlichkeits-Argumente gegen die Erwartung der Entwicklung von Reflexivität und Selbstbewußtsein in solchen Systemen zu formulieren (Abschn. 5). Den Abschluß bildet eine kurze zusammenfassende Bewertung der Überlegungen (Abschn. 6).¹⁾

2. Reflexivität

Etymologisch gesehen stammen 'Reflexivität', 'Reflexion' usw. vom Lateinischen 'reflectere', das soviel wie 'rückwärtsbiegen' 'zurückwenden' u.ä. bedeutet. Der philosophischen und psychologischen Bedeutung liegt ein Modell aus der Optik zugrunde: das des Lichtstrahls, der von einem Medium zurückgeworfen wird und gegebenenfalls in seinen Ausgangspunkt zurückkehrt. Demzufolge meint 'Reflexion' bewußtseinstheoretisch die Abwendung des Bewußtseins von den Dingen und Zusammenhängen der äußeren Welt, bei denen es sich zunächst und zumeist aufhält, und die Hinwendung und Rückkehr zu sich selbst. Diese Selbst-Zuwendung kann sich sowohl auf seine erkenntnisleistenden Prozesse wie auf die darin als Produkte zustandegebrachten Gehalte richten ("noetische" bzw. "noematische Reflexion", Wagner 1959, §§ 4,5).

Schon hier lassen sich zwei untereinander zusammenhängende Eigentümlichkeiten von Reflexivität aufzeigen, die für das Folgende bedeutsam sind. Zum einen gibt es in einem bestimmten Sinne kein 'Über-Hinaus', kein 'Außerhalb' der Reflexion: Wenn immer etwas über Reflexion ausgemacht werden soll, geschieht das in einer reflexiven Einstellung. So wie man

die Sprache selbst als Metaebene benötigt, um sie auf der Objektebene zu thematisieren (vgl. Frey 1965, p. 25 f.), so sind auch in bezug auf Reflexivität Objekt- und Metaebene nicht strikt voneinander ablösbar: Zum anderen liegt in der prinzipiellen Iterierbarkeit (Unendlichkeit) von Reflexionen die kritische Möglichkeit der Distanzierung i.S. einer Setzung des Nicht-Identisch-Seins-Mit. Verdeutlichen läßt sich dies an der von Mead herausgestellten I/me-Dialektik: Ist das Ich 'als me' qua Rollenzugehörigkeiten beschreibbar und festgelegt, so kann es sich 'als I' daraus lösen und sich neue Handlungsmöglichkeiten schaffen (Mead 1934, Kap. 22).

Bei der bisherigen Darstellung von Reflexivität ist der Aspekt ihrer Subjektfundiertheit außer acht gelassen worden, i.e. daß in der Abwendung vom Objektbezug 'ein Ich' sich 'seinem' Bewußtsein, also sich selbst zuwendet. Nun beinhaltet natürlich 'objektabgewandte' Reflexion immer auch Selbst-Zuwendung, so daß also Selbst-Bezüglichkeit und Selbst-Bewußtsein notwendig impliziert sind. Demgemäß findet sich, über die bisherigen Bestimmungen hinaus, der Ich-Rekurs auch in der von Eisler gegebenen Explikation. Danach ist "Reflexion"

"(psychologisch)... die Zurücklenkung der Aufmerksamkeit von den (primären) Außenseinsgegenständen (Naturobjekten) des Erkennens auf das psychische Erleben, auf die Bewußtseinstätigkeit und auf das erlebende oder denkende Subjekt, das Ich, auf die Formen, in denen das Subjekt das Objekt erfaßt und denkt, auf die Relationen zwischen diesen, auf die Gesetze seines theoretisch-praktischen Verhaltens, auf die Ziele und Normen seiner Tätigkeit (Selbstbesinnung)" (Eisler 1927, Bd. 3, p. 657).

Im letzten Teil dieser Bestimmung ist neben dem bisher einzig berücksichtigten theoretisch-kognitiven auch ein praktisch-handlungsbezogener Aspekt von Selbst-Bewußtsein (-Reflexion) angesprochen. Eine entsprechende Unterscheidung findet sich (auf sprachanalytischem Hintergrund) explizit bei Tugendhat.

Tugendhat unterscheidet, als Antwort auf die "Frage nach dem Phänomenbereich, auf den sich die Worte 'Selbstbewußtsein', 'conscientia', 'Sich-zusichverhalten', 'Reflexion' beziehen" (Tugendhat 1979, p. 32), zwischen einem "theoretischen" oder "epistemischen" und einem handlungsrelevanten Selbstbewußtsein. Mit ersterem sind "lediglich konstatierende" Selbst-Bezüge der Form "ich weiß, daß ich ..." gemeint, wobei dies Wissen dem Subjekt, wie bspw. bei Schmerzen, "unmittelbar" oder aber, wie etwa im Falle körperlicher Merkmale oder Charakterzüge, "mittelbar", d.h. ihm selbst prinzipiell nicht anders zugänglich als Dritten, gegeben ist (a.a.O., p. 27). Innerhalb des mittelbaren Selbstwissens gibt es einen ausgezeichneten Bereich praktischer Bedeutsamkeit: es ist das Wissen,

das sich auf die Selbstdurchsichtigkeit hinsichtlich der eigenen Handlungs- und Willensdispositionen und dafür wichtiger biographischer Ereignisse bezieht (a.a.O., p. 28). - In bezug auf das epistemische Selbstverhältnis sind Personen nach dem Substanz-Akzidenz-Modell konzipierbar. Das gilt nach Tugendhat nicht mehr, insofern sie als handelnde Wesen, die als solche in intersubjektiven Zusammenhängen stehen und die Möglichkeit zur Selbstbestimmung haben, betrachtet werden: "Als Handelnde sind wir, was wir tun und wollen; und darin haben wir schon ein Selbstverhältnis, das von anderer Art ist als das epistemische" (a.a.O., p. 29). Ein solches Selbstbewußtsein liegt, auf einer "ersten Stufe", bereits vor, indem wir "nur überhaupt etwas tun und wollen": Wir können nämlich zu eigenen oder fremden Intentionen "ja" oder "nein" sagen (a.a.O., p. 29). Dies Selbstbewußtsein kann man, auf einer "zweiten Stufe", "reflektiert" nennen, wenn man die bestimmten Intentionen zugrundeliegenden Meinungen und Überzeugungen nicht einfach als gegeben hinnimmt, sondern sie kritisch auf ihre Wahrhaftigkeit hin infragegestellt (a.a.O., p. 32).

Allen vorstehend angegebenen Facetten von Reflexivität und Selbstbewußtsein lassen sich psychologisch untersuchte Fragestellungen zuordnen. Blendet man den Aspekt des Selbstbezugs einmal aus, so finden sich wichtige Beispiele für die Thematisierung von Reflexivität etwa im Fragenkomplex der Metakognition und ihrer Entwicklung und der Bedeutsamkeit selbstreflexiven Denkens beim Problemlösen (Dörner 1976, p. 135 f.; 1979; dieser Band). - Unmittelbares epistemisches Selbst-Wissen wird methodisch im Rahmen der Selbstbeobachtung in Anspruch genommen, insofern es um die Beschreibung innerer Zustände geht (vgl. Feger & Graumann 1983). Ferner spielt es als nicht-thetisches (präreflexives) Wissen in der phänomenologisch orientierten Bewußtseinstheorie eine wichtige Rolle ("inneres Bewußtsein" bei Brentano 1874; Kap. 2; vgl. auch Wagner 1959, p. 341 f.). - Um mittelbare Selbst-Kenntnis geht es in bestimmten sozialpsychologischen Theorien der Selbstwahrnehmung bzw. -beurteilung, etwa der von Bem, wie auch bei der Frage nach dem Selbst-Konzept von Personen, die häufig praktisch-relevant gestellt wird, bspw. im Hinblick auf das Konzept der eigenen Fähigkeiten. Wissen gerade letztgenannter Art ist natürlich zentral in einer Reihe klinischer Theorienbildungen, so in der Psychoanalyse (in der ja der Selbstreflexion eine ausgezeichnete Rolle zukommt (Habermas 1973)) oder in solchen behavioralen Ansätzen, die mit dem Konzept der Selbst-Kontrolle arbeiten. - Auch die handlungsbezogenen Aspekte des Selbstverhältnisses finden sich in psychologischen Fragestellungen wieder. Entwicklungspsychologisch lassen sich etwa das Entstehen des Nein-Sagens sowie der ganze Bereich der Moralentwicklung anführen, sozialpsychologisch die "I/me"-Dialektik i.S. Meads im Rahmen des symbolischen Interaktionismus.

Die vorangehenden Ausführungen zeigen, wie verzweigt sich das Problemfeld "Reflexivität" darbietet und in welcher Vielzahl von Facetten es psychologisch relevant ist bzw. bearbeitet wird - wenn auch letzteres nur selten explizit unter diesem Titel -, ohne daß dabei ein integrierender Ansatz i.S. einer allgemeinen (psychologischen) Theorie der Reflexivität zu erkennen wäre. Es sollen daher separat drei Reflexions-Felder im Hinblick auf die Abbildbarkeit je eigentümlicher Strukturmerkmale im Rahmen des im Abschn. 3 zu präzisierenden kognitiven Paradigmas diskutiert werden: Selbstreflexion als Teilprozeß beim Problemlösen (Aspekt der System-Distanzierung, Abschn. 4.1), Selbstreflexion überhaupt (ihr Verhältnis zur Selbstreflexionstheorie; Aspekt der Abgeschlossenheit, Abschn. 4.2) und Selbstbewußtsein (Nicht-Objektivierbarkeit des 'Ich', Abschn. 4.3).

3. Explikation des Konzepts 'kognitive Psychologie'

Wie üblich soll hier 'kognitive Psychologie' zunächst ganz allgemein als Psychologie der Informationsverarbeitung, i.e. der Vorgänge der Aufnahme, Transformation, Speicherung und Verwendung von Informationen aufgefaßt werden. Betrachtet man nun den Computer nicht mehr als eine Art perfektionierte Tischrechenmaschine, sondern als allgemeines symbolverarbeitendes System (vgl. Abschn. 1), so wird unmittelbar einsichtig, daß er qua Software (also Programm) - nicht notwendig qua seiner spezifischen elektronischen und sonstigen Hardware -, ein elaboriertes 'hinreichendes' Modell für einen solchen Ansatz bietet; denn

"Computers take symbolic input, recode it, make decisions about the recoded input, make new expressions from it, store some or all of the input, and give back symbolic output. By analogy, that is most of what cognitive psychology is about" (Lachman, Lachman & Butterfield 1979, p. 99).

Hinsichtlich der 'Enge' dieser Analogie, also des Zusammenhangs zwischen Computer-Metapher und kognitiver Psychologie, lassen sich (unter Verzicht auf mögliche Abschattierungen) zwei Gruppen unterscheiden: Proponenten der 'Identitäts-Position' sehen menschliche und maschinelle Informationsverarbeitung nur als unterschiedliche Konkretisierungen desselben Prinzips an, so daß Programme zu Recht als Simulation und Abbildung menschlicher mentaler Prozesse angesehen werden (so bspw. Newell 1981; p. 72), während Proponenten der 'Nicht-Identitäts-Position' nur eine begrenzte, ein Stück weit nützliche Analogie sehen, die aber keinesfalls die geistigen Prozesse angemessen beschreibe (so Neisser 1963; 1974, p. 25 f.). Lachman et al. fassen diese Unterscheidung wie folgt zusammen:

"For some cognitive psychologists, these human properties that involve symbol manipulation are instances of the same theoretical abstraction that serve as the conceptual prototype for digital computers. At the same time, many information-processing psychologists view their theori-

zing as analogical rather than instantiatinal" (Lachman et al. 1979, p. 121; Hervorh. G. Schn.).

Die nachfolgenden Überlegungen zeigen, daß sich die Identitäts-Position bei einer bestimmten naheliegenden Präzisierung des zentralen Begriffs der Informations-Verarbeitung mit Zwangsläufigkeit ergibt. Nun ist diese Explikation aber schon derart allgemein, daß garnicht sicher ist, ob im Rahmen der Nicht-Identitäts-Position überhaupt eine entsprechende Präzisierung formuliert werden kann, die nicht in jener aufgeht: In diesem Sinne ist das Computer-Programm nicht nur ein 'hinreichendes', sondern sogar auch ein 'notwendiges' Modell für eine Informationsverarbeitungspsychologie. Diese Konsequenz wird deutlich, wenn wir, ausgehend von der Definition eines "information-processing system" bei Newell & Simon (1972, p. 20 f., p. 29 f.), zunächst ganz allgemein unter 'Verarbeitung' die als regelmäßig rekonstruierbare Zerlegung eines komplexen Ganzen in einfache elementare Teilprozesse verstehen, die ihrerseits vollständig und genau beschreibbar sind. Zur Präzisierung und Konkretisierung der zuletzt genannten Teilprozeß-Charakterisierungen wie des Aspekts der Regelmäßigkeit bietet sich das intuitive Konzept des 'Algorithmus' als eines allgemeinen, effizienten Verfahrens an, mit dem Probleme einer bestimmten Art und einer festgelegten Methode Schritt für Schritt gelöst werden (vgl. etwa Hermes 1978, Kap. 1).

In der mathematischen Logik gibt es eine Reihe unterschiedlicher Definitionsvorschläge für diesen Begriff, dessen bekanntester wohl der der Turing(maschinen)-Berechenbarkeit ist. Dabei ist eine Turingmaschine eine nur weniger elementarer Operationen, wie bspw. Zeichenerkennen, -drucken, -löschen, Rechts- und Linksverschiebung, fähige Maschine, durch die sie, entsprechend ihrem jeweiligen Programm, einen Ausgangszustand auf den sie angesetzt wird ('Input'), in einen Ergebniszustand ('Output') transformiert (vgl. Hermes 1978, p. 18/27; p. 33/38; "Algorithmtheorie" in: Enzyklopädie...Bd. 1, 1980, p. 81 f.).

Es stellt sich nun heraus, daß die verschiedenen Präzisierungen (wie bspw. Turing-Berechenbarkeit) des intuitiven (!) Konzepts des Algorithmus sich als äquivalent i.d.S. erweisen, daß die durch sie definierten Klassen berechenbarer Funktionen übereinstimmen (Hermes 1978; "Algorithmtheorie" in: Enzyklopädie...Bd. 1, p. 81/84). Über diese einzelnen Äquivalenzen hinausgehend behauptet die heute fast durchgängig akzeptierte, 1936 formulierte Churchsche These, daß in diesen Präzisierungen algorithmische Beherrschbarkeit (Berechenbarkeit) von Prozessen im intuitiven Sinne erschöpfend ausgedrückt ist: Alle Präzisierungen liefern dieselbe Klasse von Funktionen (vgl. Hermes 1978; "Churchsche These" in: Enzyklopädie... Bd. 1, p. 398 f.).

Die Churchsche These liefert sofort die Gleichheit der durch 'primitive' Turing-Maschinen und 'übliche' Computerprogramme realisierten Funktionen (Zustände), sind doch erstere nur Sonderfälle letzterer, letztere aber im intuitiven Sinne algorithmisch angelegt und damit auch durch Turing-Maschinen simulierbar (ein konkretes Beispiel eines entsprechenden Äquivalenznachweises findet sich in Newell 1981, p. 56/71). Daraus folgt aber auch die obige Notwendigkeits-Aussage bzw. Identitäts-Position: Faßt man Informationsverarbeitung als regelhaft aus Elementarprozessen zusammengesetzten Prozeß auf, der effizient bestimmbar (berechenbar) ist, so ergibt sich (notwendig) der Computer qua Programm (also qua Software) als adäquates Modell. Dies entspricht der von Wimmer und Perner formulierten These, nach denen es regulatives Ziel der Kognitionspsychologie ist, die Informationsverarbeitung Schritt für Schritt, dies algorithmisch verstanden, aufzudecken:

"Wenn dies Unterfangen im Prinzip erfolgreich sein sollte, müßte sich ... der menschliche Informationsfluß im Prinzip auf einem Allzweckcomputer simulieren lassen" (Wimmer & Perner 1979, p. 13).

Auf diesem Hintergrund stellt sich also jetzt die Frage, ob die am Ende von Abschn. 2 formulierten Reflexivitätsaspekte sich mittels algorithmisch rekonstruierbarer Systeme, wie bspw. Computerprogrammen, abbilden lassen.²⁾

4. Probleme der adäquaten Rekonstruktion von Reflexivität im Rahmen der kognitiven Psychologie

4.1. Selbstreflexion beim Problemlösen

Dörner (1979) stellt ein Modell vor, das selbstreflexives Denken als genuine Bestandteil von Problemlöseprozessen zu thematisieren gestattet. Dabei wird Selbstreflexion wie folgt bestimmt:

"Self reflection consists in the system's being able ... to make the record of its own information processing into the object of further heuristic techniques ... This means that it does not observe itself directly but only indirectly, in that it analyses the records of its own activities" (Dörner 1979, p. 106).

Die Struktur eines dazu fähigen Systems wird folgendermaßen skizziert:

- 1) Als Basis des Problemlöseprozesses stehen in der Form von Sequenzen von elementaren Teilschritten in der Informationsverarbeitung "heuristische Techniken" zur Verfügung, die in der heuristischen Struktur HS zusammengefaßt sind.
- 2) Ein "Kontrollsystem" C kann a) den jeweiligen Zustand der HS "identifizieren" und vermag b) hemmend oder bahnend auf den Prozeßablauf Einfluß zu nehmen.

3) In einem "Speicher" können die gemäß (2a) identifizierten heuristischen Prozesse festgehalten werden. Dieser Speicherinhalt kann nun seinerseits als Input für die HS (1) dienen.

Auf Grund der wegen (3) möglichen Thematisierung seiner abgeschlossenen Denkprozesse kann das System sich selbst analysieren, ohne daß es zu dem mit Reflexivität verbundenen Problem des infiniten Regreß' käme: daß die Ergebnisse von HS Input für ein weiteres heuristisches System HS_2 wären, die von HS_2 für ein drittes System HS_3 usw. (Dörner 1979, p. 105 f.).

Bei einer kritischen Betrachtung des Systems fällt zunächst auf, daß ein bestimmter Aspekt von Selbstbezüglichkeit aufgrund der nur indirekten Selbstzuwendung des Denkens ausgeschlossen ist. Gemeint ist die Möglichkeit des Mit-Bewußtseins, wie bspw. des Gewährseins des Hörens beim Hören eines Musikstücks (vgl. Brentano 1874, p. 176/180). Darüber hinaus stellen sich insbesondere zwei Fragen. Zum einen läßt sich bezweifeln, ob die Verfahren der HS; sofern sie auf verschiedenen Abstraktionsniveaus funktionieren, nicht nur formal, sondern, wie im Modell vorausgesetzt, auch konkret-operational als identisch betrachtet werden können. Kann man bspw. bei der Analyse eines Inputs in Hinblick auf gemeinsame Elemente einen konkret-identischen Prozeß annehmen, unabhängig davon, ob es sich bei diesem Input etwa um visuelles Material oder um vorangegangene abstrakte Problemlöseversuche handelt? Verschärfend kommt hinzu, daß ja durchaus Reflexionen auf einer zweiten oder noch höheren Stufe möglich sind. Würden aber die Prozesse in der HS je nach Input-Niveau unterschieden werden müssen, so ergäbe sich das oben erwähnte Regreßproblem. Die Modell-Idee ließe sich dann nur dadurch beibehalten, daß man nur eine begrenzte Zahl von Reflexionsstufen zuließe: Das aber widerspräche der prinzipiellen Iterierbarkeit reflexiver Akte.

Ein weiteres Problem ergibt sich daraus, daß das skizzierte System immer nur auf einzelne eigene Akte Bezug nehmen kann, sich aber nicht als Ganzes in einer Außenperspektive thematisieren und relativieren kann. Die Notwendigkeit einer solchen Außenperspektive und Selbst-Distanz ist bei regelgeleiteten Verhaltenssystemen dann unausweichlich, wenn deren Paradigmen (Regelsysteme) als solche thematisiert werden müssen. M.E. bedingt nun gerade das Fehlen einer solchen in einem gewissen Sinne system-relativierenden Metaperspektive gravierende Einschränkungen für eine Reihe von Problemlösetheorien. Um dies einsichtig zu machen, sollen zwei unterschiedlich komplexe Objektbereiche betrachtet werden, die niedrig-komplexe Aussagenlogik (vgl. etwa Kutschera & Breitkopf 1971, p. 17/55) und die hoch-komplexe Arithmetik der natürlichen Zahlen, verbunden mit Aussagenlogik und Prädikatenlogik erster Stufe als Ausdrucksmitteln (vgl. etwa Stegmüller 1955, p. 12/20). In beiden Fällen soll das in Frage stehende Problemlöseverhalten das Beweisen bzw. Widerlegen wahrer bzw.

falscher Aussagen in bezug auf diese Bereiche sein. Wir wollen nun annehmen, es lägen i.S. von Abschn. 3 algorithmisch formulierbare kognitive Theorien dieser beiden Verhaltensklassen vor, deren reflexiver Gehalt im Rahmen der Vorstellungen des diskutierten Modells gegeben wäre: nur eine Verarbeitungsebene und keine Außenperspektive i.S. einer umfassenden Selbst-Distanzierung. Es stellt sich nun die Frage nach der Angemessenheit dieser Theoriekonzepte. - Zur Diskussion dieser Frage müssen einige Ergebnisse der mathematischen Logik als allgemeiner Theorie formaler Systeme hinzugezogen werden.

Im Rahmen der dort thematisierten 'Formalisierung' von Gegenstandsbereichen werden auch die semantischen Begriffe 'Beweis' und 'wahre/falsche Aussage' kalkülintern, rein syntaktisch als 'Ableitung', 'Folgerung' und 'Theorem/Nicht-Theorem' abgebildet. Man erhält so das formale Modell als syntaktisches Äquivalent zum semantischen (Ausgangs-)Bereich nicht nur in bezug auf Gegenstände, Prädikate, logische Operationen usw., sondern auch in bezug auf den Beweisbegriff. Das zentrale Problem ist nun, 'wie weit' die Äquivalenz zwischen formalem Modell und semantischem Bereich im Hinblick auf das Beweisen geht. Dazu gehört zunächst die Frage nach der "Korrektheit" des syntaktisch eingeführten Ableitungsbegriffs: Liefert jedes nach den Regeln des formalen Modells abgeleitete Theorem, also jeder 'syntaktisch wahre' Satz, semantisch interpretiert auch eine wahre, im üblichen Sinne mathematisch beweisbare Aussage? Diese Frage läßt sich bei allen 'vernünftigen' Formulierungen gleichsam als sine qua non der Formalisierung positiv beantworten. In unserem Kontext ist die umgekehrte Fragerichtung entscheidend, die nach der "Vollständigkeit" des Kalküls: Sind die den semantisch wahren Aussagen entsprechenden formalen Zeichenreihen sämtlich auch Theoreme, d.h. innerhalb des formalen Systems rein kalkülgemäß (syntaktisch) generierbar? Hierbei geht es also darum, ob ein formales Modell hinsichtlich des Beweizens mit seinem präzisen, fixierten Kalkül die ganze Fülle der Möglichkeiten des semantisch interpretierten Bereichs ausschöpfen kann.

Es zeigt sich nun, daß die Antwort auf die Vollständigkeitsfrage von der Komplexität des Gegenstandsbereichs abhängt. Sie fällt positiv aus (u.a.) in bezug auf die oben erwähnte Aussagenlogik (vgl. etwa Kutschera & Breitkopf 1971, p. 67/70). Dagegen fällt sie für die Arithmetik i.o.S. negativ aus, sofern man sie als konsistent (widerspruchsfrei) voraussetzt, wie erstmalig 1931 von Gödel bewiesen wurde: Es gibt arithmetisch wahre Sätze (semantische Ebene), die nicht formal-kalkülgemäß ableitbar sind (syntaktische Ebene) (vgl. etwa Nagel & Newman 1958, Stegmüller 1959). Der entsprechende Beweis kann auch explizit algorithmusbezogen formuliert werden, so daß die angegebene Unvollständigkeitsaussage gleichbedeutend damit ist, daß kein Algorithmus existiert, mit dem genau die wahren

arithmetischen Aussagen ableitbar sind (Hermes 1978, p. 176/187; Hofstadter 1979, Kap. 14).

Was implizieren diese Ergebnisse nun für unsere Fragestellung? Zum einen zeigt sich, daß das kognitive Ideal einer algorithmisch formulierten Theorie des Beweisens (als einem speziellen Problemlöseverhalten) hinsichtlich des Gegenstandsbereichs 'Aussagenlogik' nicht in Schwierigkeiten auf Grund der Vollständigkeitsproblematik gerät. Dies passiert aber in bezug auf den Bereich 'Arithmetik'; denn eine algorithmische Theorie kann prinzipiell nicht alle als wahr einsehbaren (beweisbaren) Aussagen in diesem Bereich als solche einsichtig machen: In diesem Sinne entzieht sich das Beweisen hier einer endlichen kalkülhaften (algorithmischen) Beschreibbarkeit.

Für die Reflexivitätsproblematik speziell ist interessant, daß zum Nachweis der Unvollständigkeit Sätze konstruiert werden können, die von Außen, aus einer distanzierenden Metaperspektive, unmittelbar als wahr einsichtig sind, während sie systemintern weder folgerbar noch widerlegbar sind. Ohne auf eine solche Konstruktion hier eingehen zu können, sei nur erwähnt, daß es bspw. möglich ist, eine Formel G zu definieren, die, metamathematisch betrachtet, reflexiv etwas über sich selbst zum Ausdruck bringt, nämlich: "G ist kein Theorem" (i.S. formaler Ableitbarkeit). Wäre G formal ableitbar, so ergäbe sich qua Definition von G ein Widerspruch - im Widerspruch zur vorausgesetzten Konsistenz der Arithmetik. Folglich ist G nicht formal ableitbar, d.h. kein Theorem. Damit ist G aber metamathematisch, 'von außen' betrachtet, wahr, denn die genannte Eigenschaft ist ja gerade G's Inhalt. G ist also ein Beispiel für eine semantisch wahre (Außenperspektive), nicht aber formal ableitbare Aussage (Innenperspektive). - Betrachten wir noch die Negation von G. Da G wahr ist, ist nicht-G falsch und folglich wie G ebenfalls nicht formal ableitbar, da ja auf Grund der Korrektheit des Kalküls (s.o.) überhaupt nur semantisch wahre Aussagen formal ableitbar sind. Es ist demnach nur 'von Außen', nicht kalkülintern entscheidbar, ob G oder nicht-G wahr ist (vgl. Hofstadter 1979, p. 449).

Es zeigt sich also, daß sich in jeder algorithmisch angelegten Beweisprozedur ein Satz angeben läßt, der von außerhalb des Systems, in einer reflexiven Distanzierung von seinem Verfahrensrepertoire, unmittelbar als wahr einzusehen ist, ohne daß dies aber mit den Mitteln des Systems selbst möglich wäre.

4.2. Selbstreflexion überhaupt

Nach dem Aspekt der Selbst-Distanzierung wollen wir im folgenden auf den ebenfalls bereits im Abschn. 1 erwähnten Aspekt der 'Abgeschlossenheit' von Reflexion zurückkommen: Jede Reflexions-Theorie thematisiert wegen ih-

rer eigenen Reflexionshaltigkeit auch sich selber. Dabei soll "Reflexion" hier phänomenologisch i.S. der Husserlschen Darstellung in den "Ideen zu einer reinen Phänomenologie und Phänomenologischen Philosophie" (Bd.1, 1913) gefaßt werden.

Husserl kennzeichnet Bewußtsein durch seine "Intentionalität", die er am Cartesischen "cogito", dem "Ich denke", "Ich nehme wahr", "Ich erinnere mich", "Ich urteile, fühle, begehre" usw. (Husserl 1913, p. 75) expliziert: Jede entsprechende "cogitatio" (Bewußtseinserlebnis) erweist sich als "unabtrennbar" von ihrem "cogitatum", dem Gedachten, Wahrgenommenen usw. (a.a.O., p. 77), d.h. "ein Wahrnehmen ist Wahrnehmen von etwas, etwa einem Ding; ein Urteilen ist Urteilen von einem Sachverhalt ... usw. Handeln geht auf Handlung. Tun auf Tat. Liebe auf Geliebtes ... usw." (a.a.O., p. 204). Am Beispiel des Wahrnehmens läßt sich dabei folgender Aspekt gewinnen: Nicht nur haben wir in unserem Wahrnehmungsfeld das Objekt selbst, auf das wir bezogen sind - Husserl spricht hier vom "Modus der aktuellen Zuwendung" oder "Aktualität" (a.a.O., p. 78 f.)-, sondern um dieses herum zugleich auch einen "Hof von Hintergrundsanschauungen", von denen nur ein "implizites, bloß potentielles (Bewußtsein)" vorliegt - diese Erlebnisse stehen im "Modus der Inaktualität" (a.a.O., p. 77 f.).

Ein ähnliches Verhältnis von Hervorgehobenheit und Hintergrund besteht im sich vollziehenden "cogito" selbst: "Im cogito lebend, haben wir die cogitatio selbst nicht aktuell bewußt als intentionales Objekt", wenn auch "zu ihrem Wesen" die "prinzipielle Möglichkeit" dazu gehört (a.a.O., p. 184). Die 'Aufschließung' der cogitationes wird nun im Akt der reflexiven Selbstzuwendung geleistet. Dabei ist zu beachten, daß gegenüber den Dingen der Außenwelt die cogitationes in besonderer Weise ausgezeichnet sind: Während jene, sofern sie sich nicht im Hintergrund des Wahrnehmungsfelds befinden, nicht geradewegs in den Blick genommen werden können, sondern nur über "motivierete Wahrnehmungsreihen" in einen "Wahrnehmungszusammenhang" rücken, sind diese dem Ich, dem sie zugehören, in einer einfachen, "schlicht-erfassenden", "gewahrenden" (a.a.O., p. 84) In-Beachtung-Nahme zugänglich (a.a.O., p. 104 f.). Das zentral Wichtige an dieser "reflektiven Blickzuwendung" (a.a.O., p. 84) oder "Reflexion" (a.a.O., p. 104) ist demnach, daß durch sie die Gesamtheit des Bewußtseinslebens für das Ich zum Objekt werden kann (a.a.O., p. 178). Damit wird die Bedeutsamkeit von Reflexivität auch i.S. der oben erwähnten 'Abgeschlossenheit' im Rahmen dieser Bewußtseinstheorie deutlich: "(Reflexion) ist ... der Titel der Bewußtseinsmethode für die Erkenntnis von Bewußtsein überhaupt" (a.a.O., p. 181).

Natürlich können Reflexionen als Erlebnisse wieder zum Gegenstand der Reflexion werden, ohne daß diese Reihe ausschöpfbar wäre: "Die Reflexionen

sind abermals Erlebnisse und können als solche Substrate neuer Reflexionen werden und so in infinitum, in prinzipieller Allgemeinheit" (a.a. O., p. 178).

Es stellt sich nun die Frage, ob Reflexivität, wie sie vorstehend skizziert wurde, i.S. des in Abschn. 3 entwickelten Verständnisses kognitiver Psychologie strukturell adäquat algorithmisch rekonstruierbar ist. Eine solche Theorie wäre ein algorithmisches System, für das auf Grund des In-Eins-Seins von Objekt- und Metaebene in der Reflexion, ihrer (reflexiven) Abgeschlossenheit, gelten würde, daß alle wahren Aussagen über das System im System selbst ausdrückbar, d.h. rein syntaktisch gewinnbar wären: Eine algorithmisch formulierte Theorie der Reflexion wäre i. S. von Abschn. 4.1 vollständig. Nun gilt nach Rosser, daß widerspruchsfreie Obertheorien, die die erweiterte Arithmetik umfassen, wie schon die Arithmetik selbst unvollständig sind (Hermes 1978, p. 178). Demnach ist zu fragen, ob die Arithmetik strukturell als Teil einer (als widerspruchsfrei anzunehmenden) Reflexionstheorie aufgefaßt werden kann. Dies kann wie folgt begründet werden.

Zunächst einmal wären die Operationen der der Arithmetik zugeordneten Ausdrucksmittel Aussagen- und Prädikatenlogik erster Stufe, wie bspw. Konjunktionen, Negationen, Quantifizierung, Inferenz, als elementare Ausdrucks-, Verknüpfungs- und Ableitungsmöglichkeiten sicherlich in einer solchen Theorie zu repräsentieren. Andererseits rekuriert aber auch die Peanosche Charakterisierung der natürlichen Zahlen (im folgenden in Anlehnung an Hermes (1963, p. 13) formuliert) wesentlich auf reflexive Operationen: 1) Man beginnt mit einer cogitatio (1. Peano-Axiom: 0 ist eine natürliche Zahl), die 2) zum Ausgangspunkt einer sukzessiv wiederholten je neuen Reflexion gemacht wird (Nachfolgerbildung: $0, N(0), N(N(0))$ etc.), wobei jede solche Reflexion wieder als cogitatio aufzufassen ist (2. P.-A.: jeder Nachfolger jeder natürlichen Zahl ist wieder eine nat. Z.), die 3) stets von der Ausgangscogitatio verschieden ist (3. P.-A.: der Nachfolger einer nat.Z. ist stets von 0 verschieden), und die 4) dann von einer zweiten Reflexion verschieden ist, wenn die vorausliegenden cogitationes nicht identisch sind (4. P.-A.: die Nachfolger verschiedener nat.Z. sind wieder verschieden). In gewisser Weise läßt sich auch 5) das Induktionsaxiom (5. P.-A.: trifft eine Eigenschaft auf 0 zu und stets auch auf den Nachfolger einer nat.Z. x zu, sofern sie nur auf x selbst zutrifft, so trifft sie auf alle nat.Z. zu) reflexiv interpretieren, nämlich als Zurückwendung von den Analysen eines Einzelglieds (nämlich der 0) und denen aller 1-gliedrigen Übergänge ($x \rightarrow x+1$) auf die Gesamtheit aller Reihenglieder selbst.

Betrachtet man auf Grund dieser Überlegungen die Arithmetik als Teil einer (wesentlich) umfassenderen Reflexionstheorie, so ist nach Obigem jede Formalisierung einer solchen Reflexionstheorie unvollständig i.d.S., daß es nicht formal ableitbare wahre Aussagen über die Reflexion gibt. Diese gehören andererseits aber wegen des Zusammenfalls von Objekt- und Metaebene zur Reflexionstheorie selbst: Es kann also keine vollständige algorithmische (formale) Reflexionstheorie geben. Theorien dieser Art können prinzipiell nicht den vollen Gehalt von Reflexivität ausschöpfen. Das Abgeschlossenheits-Argument ist von Frey (1965, 1975, 1980) adaptiert, der es auf einer anderen Ebene verwendet. Nach ihm findet sich Bewußtseinsreflexivität in sprachlicher Reflexivität abgebildet. Da nun in der Umgangssprache als unhintergehbarem Verständigungsmedium überhaupt Objekt- und Metaebene 'zusammenfallen', ergibt sich (wenn auch technisch anders begründet) dasselbe Resultat.

4.3. Selbstbewußtsein

Im vorangehenden Abschn. 4.2 wurde beiläufig auf die besondere Rolle von Reflexion im Rahmen der Subjektivität hingewiesen: durch sie wird das Bewußtseinsleben für das Ich erschlossen (Husserl 1913, p. 178)³⁾. Daran anknüpfend stellt sich die Frage nach der Notwendigkeit der Thematisierung der Ich-Perspektive. Dementsprechend soll zunächst im Anschluß an H. Schmitz dargelegt werden, daß tatsächlich die 'Erste-Person-Perspektive', wie sie sich sprachlich in Wendungen der Form 'ich habe ...', 'ich bin ...' oder '... für mich' u.ä. ausdrückt, nicht auf eine objektive Außenbeschreibung reduzierbar ist, die ohne Rekurs auf die Ich-Sphäre auskommt, also eine Dritte-Person-Perspektive⁴⁾.

Schmitz (1969) führt den Nachweis der Unverzichtbarkeit von Erste-Person-Ausdrücken mit Hilfe sprachanalytischer und phänomenologischer Argumente. Im Zentrum seiner Überlegungen stehen die sog. "strikt-ichbezogenen" Ausdrücke (a.a.O., p. 40). Dabei werden "ichbezogen" solche Ausdrücke genannt, deren Verwendung außerhalb bloßer Anführung "gemäß den sprachlichen Regeln höchstens während des Vollzugs eigenen Selbstbewußtseins möglich (ist)" (a.a.O., p. 36), bspw. "heute", "hier", "ich", "mein" (a.a.O., p. 36 f., p. 40). Die Ichbezogenheit eines Ausdrucks ist "strikt", wenn bei seiner Verwendung durch unterschiedliche Sprecher (zur gleichen Zeit und am gleichen Ort) trotz äußerlich gleicher Form stets unterschiedliche Sachverhalte ausgesprochen werden; dies gilt bspw. für "ich", nicht aber für "hier", wie die Beispielssätze "Ich bin traurig" bzw. "Hier ist es schön" zeigen (a.a.O., p. 39 f.). Für Schmitz "(sind die) einzigen strikt ichbezogenen deutschen Ausdrücke ... die Personal- und Possessivpronomen der ersten Person des Singulars sowie die von jenen Personal-Pronomina regierten Verbformen derselben grammati-

schen Kategorie" (a.a.O., p. 40).

Es stellt sich nun die Frage, ob die mit Hilfe strikt ichbezogener Ausdrücke formulierbaren Sachverhalte in eine reduzierte Sprache S', die weder explizit noch implizit auf sie rekurriert, sachverhaltsgetreu übersetzbar sind (a.a.O., p. 47 f.). S' wäre "eine gewissermaßen vollkommen objektive Sprache, in der sich alles das und nur das beschreiben läßt, was überhaupt beschrieben werden kann, ohne daß der Sprecher etwas als sich selber meint und anerkennt" (a.a.O., p. 49, Hervorh.G.Schn.).

Woran Übersetzungsversuche aus der vollen ich-haltigen Sprache S in die reduzierte S' scheitern, zeigt Schmitz am Beispielsatz: (1) "Ich bin traurig", insofern dieser als wahre Selbst-Aussage gemeint ist (a.a.O., p. 47). Wird "Ich" in S' bspw. durch den Eigennamen ersetzt, so ist der resultierende Satz (2) "Herrmann Schmitz ist traurig" (a.a.O., p. 49) zwar dem Buchautor verständlich, da ihm H. Schmitz durch bestimmte Kennzeichnungen bekannt ist, doch entsteht eine Sinndifferenz gegenüber (1) dadurch, daß er sich ja lt. Definition von S' nicht mit dem Referenten des Eigennamens identifizieren darf (s.o., Zit.): Nun bringt aber (1) gerade sein persönliches Betroffensein zum Ausdruck, und zwar nicht als beiläufiges Mitgegebensein, sondern als wesentliche Sinn-Nuance. "Diese Nuance fehlt aber im Sinn des Satzes ... (2) der verarmten Sprache, da ich laut Voraussetzung beim Verständnis dieses Satzes die Identifizierung mit mir unterlassen habe ..." (a.a.O., p. 49 f.). Dieser Sinngehalt ist auch nicht durch irgendwelche nicht ichbezogene Operationalisierungen persönlichen Engagements ausschöpfbar, da stets das Identifikationsmoment entfällt (a.a.O., p. 50 f.). Es folgt also, daß "(ich) ... den Gedanken, daß etwas mit mir identisch ist, nicht anders als so fassen (kann), daß eine vollständige Formulierung dieses Gedankens einen strikt ichbezogenen Ausdruck ... enthalten würde. Ohne diesen Weg einzuschlagen, kann ich zwar immer noch mich selbst ... aber nicht etwas als mich selbst meinen" (a.a.O., p. 50, Hervorh.G.Schn.).

Auf dem Hintergrund der dargelegten Nicht-Reduzierbarkeit der Ersten-Person-Perspektive (EPP) auf eine Dritte-Person-Perspektive (DPP) sollen nun zwei Fragen diskutiert werden: Ist zum einen der Übergang von Beschreibungen in der DPP, die tatsächlich einen selbst betreffen, zur EPP im Rahmen der kognitiven Psychologie, also algorithmisch, rekonstruierbar?, und wie steht es zum anderen mit der programmbezogenen Realisierung der EPP überhaupt?

Daß jemand (in der DPP) auf sich Bezug nehmen kann, ohne daß er (in der EPP) damit sich als sich selbst meint, sieht man am Fall des Ödipus, wenn er als König von Theben Diener ausschickt, die den Mörder des Laios finden sollen: Da er selbst dieser Mörder ist, schickt er in der Tat

seine Diener aus, ihn selber gefangenzunehmen, ohne daß er dies aber mit seinem Befehl intendierte; er bezieht sich also nur objektiv, nicht aber selbst-bewußt auf sich (Baker 1981, p. 158). Alltäglichere Beispiele ergeben sich, wenn wir etwa Beschreibungen über oder Aufforderungen an Klassen von Personen formulieren, zu denen wir selbst faktisch auch gehören, ohne daß wir aber diese Zugehörigkeit in der EPP realisieren, i.e. auch uns selbst meinen würden. Systematisch gehören in diesen Kontext projektive psychodynamische Abwehrmechanismen, insofern man sie als Transformationen unbewußter Inhalte der Form 'Selbst-Handlung-Objekt' in bewußte der Form 'Anderer-Handlung⁽¹⁾-Objekt⁽¹⁾' auffassen kann (Suppes & Warren 1975, p. 408 f.).

Offensichtlich gibt es nun einerseits Übergänge von der DPP zur EPP, wie sich auch andererseits die DPP als 'übergangsresistent' erweisen kann. So akzeptiert bspw. Ödipus schließlich, daß er selbst Laios erschlagen hat. Auf der anderen Seite können aber entsprechend ausgearbeitete Wahnsysteme durchaus gegen jeden von außen herangetragenen Falsifikationsversuch und damit gegen jeden Versuch, eine EPP herzustellen, immun sein. Entscheidend ist nun im vorliegenden Problemzusammenhang bzgl. der Programmierbarkeit des Selbstbewußtseins, ob der Übergang von der DPP in die EPP als allgemein regelhaft angenommen werden kann. Unter der Voraussetzung einer bereits gegebenen EPP impliziert dies die folgende Frage: Es sei ein bestimmtes DPP-Wissen über einen selbst vorausgesetzt; läßt sich dann mittels eines spezifizierbaren Informationszuwachses ein Übergang in die EPP regelhaft herbeiführen?

Schon auf Grund der o.a. Beispiele läßt sich vermuten, daß dies eher unwahrscheinlich ist. Baker formuliert schärfer:

"... there is no way to specify how much information about oneself in the third-person is sufficient to lead one to first-person belief; the variety of conditions under which we make the leap to first-person belief is so extraordinary that it is futile to look even for rules to govern all the appropriate transformations of third-person sentences to first-person sentences" (Baker 1981, p. 161).

Sicherlich ist der Status dieses Arguments anders als der der vorangehenden Abschn. 4.1/4.2: Es ist logisch gesehen weniger strikt und nimmt Rekurs auf unser nicht systematisiertes Erfahrungswissen. Der tiefere Grund dafür liegt wohl darin, daß hier der kognitive Aspekt des Wissens-um-einen-selber nicht losgelöst vom Bereich der Affektivität gesehen werden kann, wie es schon die Beispiele andeuten. Eine striktere Begründung für die wahrscheinlich nicht als regelhaft anzunehmenden DPP/EPP-Übergänge würde also eine komplexe Theorie des Zusammenhangs dieser Bereiche unter der Einheitsperspektive 'Subjektivität' erfordern.

Bisher ist hypothetisch vorausgesetzt worden, daß die Unterscheidung zwischen EPP und DPP adäquat in einem Programm abgebildet werden könne. Wie steht es aber mit dieser Voraussetzung selbst: Ist die DPP/EPP-Differenz und damit Selbstbewußtsein programmierbar? Daß dies in einem endlichen Programm nicht möglich ist, kann wie folgt eingesehen werden: Man braucht dazu nur die endlich vielen Ausdrücke, mit denen in diesem Programm auf es selbst Bezug genommen wird, durchzusehen und alle die auszusondern, die für die EPP verwendet werden. Diese Ausdrücke werden dann bspw. durch Eigennamen oder Kennzeichnungen ersetzt, die nicht in der endlichen Menge der bereits verwendeten DPP-Ausdrücke vorkommen. Durch diese Substitutionen wird an der Funktionsweise des Programms nichts geändert, es bleibt also mit sich identisch. Andererseits zeigen sie aber, daß sein 'Erste-Person-Bezug' zu sich selbst vollständig in einer objektivierten Außen-sprache (DPP) beschreibbar ist, während dies nach den obigen Analysen im Falle des menschlichen Selbst-Bewußtseins nicht möglich ist. Wir schließen also, daß Selbstbewußtsein nicht mittels eines (endlichen) Programms darstellbar ist^{5,6}.

5. Reflexivität und Selbstbewußtsein bei lernenden interaktiven Systemen

Im Zentrum der bisherigen Überlegungen stand die Frage nach der theoretischen Rekonstruierbarkeit von Reflexivität und Selbstbewußtsein im Rahmen der kognitiven Psychologie. Man kann nun versuchen, die dargestellten negativen Befunde auf das darin gelegene statisch-restriktive Moment zurückzuführen und im Gegenzug auf die Lernfähigkeit und Nicht-Identität entsprechender Systeme zu verweisen, wie dies Boden tut:

"... Goedel's theorem applies only to closed systems, wherein all the axioms and inference rules are fixed. If a program (like a person) is capable of learning new rules and axioms, by communicating with a teacher or with the outside world and extending its internal representations accordingly, then something that was undecidable yesterday may be decidable today" (Boden 1977, p. 434). -

Ebenfalls im Zusammenhang mit einer Diskussion des Gödel-Arguments weist Dennett auf die Differenz zwischen einer theoretisch konzipierten und einer aktualisierten Turingmaschine hin, insofern letztere über ihren Entwurf 'hinausgehe':

"... Turing machine specifications say very little about the machine they specify. Characterizing something as a TM_k ascribes certain capacities to it and puts certain limitations on these capacities, but says nothing about other features or capacities of the thing ... (Nothing) concrete could be just a particular Turing machine, and any concrete realization of any Turing machine can in principle have capacities under

one interpretation denied it under another" (Dennett 1979d, p. 263 f.). Offensichtlich gibt es jetzt zwei Möglichkeiten: Entweder kann ein solches komplexes interaktiv-lernendes System, das auf Grund von aus der Außenwelt erhaltenen Informationen sein Verhalten nach bestimmten Kriterien verbessert, wobei die endgültigen Verhaltensweisen nicht voraus-sagbar sein müssen, hinsichtlich seiner Reflexionsfähigkeiten und seines Selbstbewußtseins im Rahmen der kognitiven Psychologie (sensu Abschn. 3) rekonstruiert werden oder dies erweist sich als unmöglich. Im ersten Fall ergeben sich (zumindest) die in Abschn. 4 herausgestellten Strukturdif-ferenzen zu den entsprechenden menschlichen Phänomenen, so daß keine si-mulative Isomorphie vorliegt: Das Problem ist somit auf einen bereits bekannten Fall zurückgeführt.

Wie aber steht es mit der zumindest implizit in den Humanbereich weisen- den Alternative? Ich sehe nicht, wie hier ein strikt negatives Argument aussehen könnte⁷⁾, meine aber, daß bestimmte Befunde auf derart hohe untere Grenzbedingungen i.S. einer maschineninternen wie umweltbezogenen sine qua non-Komplexität für die autochthone (onto- oder phylogenetische) Entwicklung von Reflexivität und Selbstbewußtsein hinweisen, daß zumin- dest ein Schein von romantischer shelley-ficial Science Fiction auf eine entsprechende fiktive Technologie des arti-ficial Reflexiven fiele. Be- trachten wir zunächst die hier rein funktional (!) verstandene Hardware.

Eine solch elementare reflexive Erfahrung, wie sie im spiegelbildlichen Erkennen eigener Körperteile als eigener vorliegt, tritt nach entwick- lungpsychologischen Befunden erst ab ca. 15 Monaten auf, setzt also be- trächtliche Reifungs- und Lernschritte voraus ("self recognition", Lewis & Brooks-Gunn 1979, p. 29/68). Analoge Selbst-Erkennungsversuche (vgl. bspw. Gallup 1977) wurden mit den unterschiedlichsten Tierarten durchgeführt, wobei die Selbst/Andere-Unterscheidung erst bei den höhe- ren Primaten (Schimpansen, Orang-Utans), nicht aber bereits bei niedere- ren (wie Rhesusaffen oder Gibbons) zu erkennen ist (Lewis & Brooks-Gunn 1979, p. 18 ff.), obwohl diese den ersteren ansonsten in einer Vielzahl kognitiver Leistungen nicht nachstehen (Gallup 1977, p. 333). Gallup ver- mutet, daß die Selbst-Identifizierungsleistung eine bestimmte kortikale Komplexität, die also erst bei den höheren Primaten vorläge, voraussetzt: "We are a long way from being able to specify a neurological basis for the sense of identity. However, self-recognition may represent an emer- gent phenomenon that only occurs once a species acquires a certain number of cortical neurons with sufficiently complex interconnections" (Gallup 1977, p. 334).

Darüber hinaus weisen Ergebnisse in Hinblick auf die Sprachfähigkeiten von Menschenaffen auf eine große Kluft zum Humanpotential hin. So kommen

Ristau und Robbins in einem breit angelegten Überblick zu dem Schluß: "The apes do sometimes use signs, chips, or lexigrams as symbols, although precisely how similar their use of words is to man's is yet to be determined ... There is, however, no convincing evidence that apes' utterances are grammatical ... Compared to young children, apes acquire signs much more slowly and sequence them within an utterance very differently. It is difficult to conclude that the productions of "linguistic-like" apes bear more than a rudimentary similarity to human language" (Ristau & Robbins 1982, p. 247). Die Befunde speziell zur sprachlichen Reflexivität als "Fähigkeit, über das Kommunikationssystem selbst zu kommunizieren" (a.a.O., p. 148), sind "äquivok" (a.a.O., p. 231). Bestenfalls ließe sich, bei Hinterrückstellung methodologischer Probleme (a.a.O., p. 191), behaupten, daß mehrfach gewisse Metazeichen, wie bspw. "Name von ...", gelernt wurden (a.a.O., p. 232), was natürlich nur einen Bruchteil sprachlich möglicher Reflexivität ausmacht (vgl. Frey 1980, p. 19/27).

Aufgrund der angeführten Befunde vermuten wir, daß die Entwicklung von Reflexivität und Selbstbewußtsein bei lernenden Systemen eine Hardware voraussetzen würde, deren funktionelle Komplexität der des menschlichen ZNS nahekommen müßte (Hardware-Argument).

Die Bedeutsamkeit sozialer Interaktionen für die Entwicklung des Selbst-Erkennens geht aus einem weiteren Teil der bereits erwähnten Versuche von Gallup (1977) hervor.

Beim Vergleich zwischen solchen Schimpansen, die in einer Gemeinschaft aufgewachsen waren, mit isoliert aufgewachsenen zeigten sich nämlich nur bei ersteren eindeutige Hinweise auf Selbst-Erkennen (a.a.O., p. 336). Gallup interpretiert diesen Befund auf dem Hintergrund des Cooley-Meadschen Widerspiegelungskonzepts (a.a.O., p. 335 ff.), das 'sich-sehen' als 'sich mit den Augen der anderen sehen' interpretiert und somit Selbstbewußtsein interaktiv fundiert, wobei im menschlichen Bereich der Sprache eine ausgezeichnete Rolle zukommt (Mead 1934). Entwicklungspsychologisch konstatieren Lewis & Brooks-Gunn: "Interaction between self and other is important, because it is from this interaction that both self and other are formed" (Lewis & Brooks-Gunn 1979, p. 257).

Von welcher Bedeutung eine nicht-deprivierte 'sprachlich-reflexive' Umwelt ist, zeigen die Fälle vollständig isoliert aufgewachsener Kinder, wie bspw. der "wilden Kinder" (Malson, Itard & Mannoni, 1972) oder des Mädchens "Genie", das 13 Jahre in Isolation gehalten wurde und ohne sprachliche Kommunikation aufwuchs (hier nach der Darstellung von Eccles). Eccles schließt:

"Zusammenfassend hat die tragische inhaltliche Deprivation aller sprachlichen Inputs die fundamentale Rolle der Sprache bei der Erschaffung

einer menschlichen Person mit kognitiven und kreativen Fähigkeiten gezeigt. Das deprivierte Gehirn war noch immer in der Lage, einige seiner latenten Fähigkeiten wiederzugewinnen. ... Jedoch Genies stark verzögerte Sprachentwicklung war mit vielen Schwierigkeiten behaftet, und sie ist immer noch sehr inadäquat" (Eccles in: Popper & Eccles 1982; p. 375 f.).

Nach diesen Befunden vermuten wir, daß für die Entwicklung von Reflexivität und Selbstbewußtsein ein nicht-depriviertes, sprachlich vermitteltes (und insofern humanoides) Interaktionsgefüge vorauszusetzen ist (Interaktions-Argument).

Welche Folgerungen ergeben sich aus dem zuletzt Gesagten für die Maschinen-Ontogenese? Zunächst einmal könnte man den Vorschlag erwägen, die Maschine in eine entsprechende optimal programmierte Reaktionsumwelt als 'Entwicklungs-Setting' hineinzustellen. Diese Möglichkeit entfällt aber als *petitio principii*, da das Komplexitätsniveau der Umwelt wegen ihres sprachlich-reflexiven Gehalts schon das der Maschine vorwegnahme, das diese ja als erstes mechanisches Exemplar überhaupt allererst erreichen soll. (Ferner könnte man auf die 'programmierbarkeits-negativen' Befunde von Abschn. 4 verweisen). Es scheint also, daß einzig Menschen als geeigneter 'Umgang' für die 'junge Maschine' in Frage kämen, wobei dieser Umgang sich vom analogen humanen Interaktionsgeschehen hinsichtlich seiner Qualität nicht deprivativ unterscheiden sollte. Um nicht in krude Malfiction abzugleiten, sei nur auf den Reziprozitätsaspekt menschlichen Handelns verwiesen: Die anzuzielende Interaktionalität erweist sich demnach zwar als notwendig, kaum aber als herstellbar. - Ein zweites, realitätsnäheres Argument: Wie erwähnt müssen der Maschine exakte Kriterien zur Beurteilung von Verhaltensweisen bzw. -resultaten als 'besser' oder 'schlechter' zur Verfügung stehen. Dazu müßte aber die Entwicklungspsychologie ein möglichst vollständiges zu Reflexivität und Selbstbewußtsein führendes Verhaltensrepertoire einschließlich Bewertungsstandards aufzählen können. Ob eine solche Fragestellung empirisch sinnvoll ist, kann füglich bezweifelt werden, da es sich nach Obigem hier um langfristige, komplex-verschränkte biologisch-physiologische und soziale Prozesse handelt. Betrachtet man zudem die extreme Vielfalt von Ausgangsbedingungen, Verhaltensweisen etc. auf der Ebene der konkreten Interaktion, so scheint die Idee eines präzisen (da qua Programm formulierbaren!) Bewertungsregelsystems in diesem Zusammenhang unangemessen zu sein.

Summa summarum lassen diese Überlegungen einen nicht einmal approximativ aufhebbaren Hiatus zwischen Ziel und gradueller Zielverwirklichung vermuten: Es scheint, als ob zur definitiven Überwindung dieses Hiatus ein nicht mehr mechanisch-algorithmisch rekonstruierbarer fulgurativer Akt (sensu Lorenz 1973, p. 48 ff.) als Creator reflexiven und selbstbewußten

Maschinenseins vonnöten sei (Fulgurations-Argument).

Nimmt man das Hardware-Argument und das aus dem Interaktions-Argument gefolgerte Fulgurations-Argument zusammen, so legt sich im Gedenken an die romantisch-literarische Ahnfrau des Sujets die Mary W-Shelley-Hypothese, mechanisch-funktionelle Version, nahe: 'Der Bau eines Automaten, der Reflexivität und Selbstbewußtsein entwickelt, gehört in funktionaler Hinsicht zur Gruppe der Frankenstein-Operationen'.

Die bisherigen Überlegungen beziehen sich nur auf das Problem der Ontogenese maschinell realisierten Selbstbewußtseins. Phylogenetisch-evolutionstheoretisch eröffnen sich ganz neue Möglichkeiten, die limitativen Argumente den Atem rauben: Hier hilft nur noch die Empirie. Schicken wir also zu gegebener Zeit sapiens maior-Kandidaten auf den Mond oder Odysseen bspw. zum Jupiter und warten, eingedenk unserer eigenen Geschichte als Menschheit, ab, warten auf die unverzichtbare Fulguration, die dann am Ende des Quartärs, nach dem Übergang zum Leben überhaupt und später vom tierischen zum menschlichen Leben (Lorenz 1973, p. 227 f.), jenen dritten großen Hiatus schafft, auf Grund dessen wir gegen Ende des Quintärs sodann als ein verkümmender Seitenzweig am Deszendensbaum der Maioriden möglicherweise als ein Objekt ihrer Neugier entdeckt werden bis hin zur Alternative 'stirb oder paß dich an', die in Vorzeiten einst Kafka einen unserer Minoriden in seinem Selbstwerdungsbericht entfalten ließ.

6. Schlußbemerkung

Es sollte zunächst noch einmal festgehalten werden, daß es im vorliegenden Kontext, wie in Abschn. 1 dargelegt, nicht um die Analyse explizierter Modelle von Subjektivität und Bewußtsein gehen sollte - Dennett (1979c, p. 149 f.) wie auch Norman (1981a, p. 279 f.) konstatieren kritisch eine entsprechende Enthaltensamkeit der kognitiven Psychologie -, sondern um die metatheoretische Frage nach Grenzen des Computer(Programm)-Paradigmas im Hinblick auf Reflexivität und Selbstbewußtsein als zentralen Aspekten von Subjektivität. Nun ist aber Subjektivität Thema bzw. regulative Zielidee innerhalb der kognitiven Psychologie: So sind nach Boden (1981a; 1981b, p. 37/49) strikt subjekthafte Begriffe, wie bspw. Intentionalität, hier aufhebbar und gehören zum theoretischen Vokabular; nach Newell muß ein System, das definitiv als "mind-like" qualifizierbar sein soll, u.a. der folgenden Forderung genügen:

"... Exhibit self-awareness and a sense of self" (Newell 1981, p. 41).

Damit gewinnt also unsere Fragestellung schon vor dem Problembereich "Subjektivität" im Rahmen der kognitiven Psychologie und nicht erst vor dem unspezifischeren Hintergrund der 'no limits'-Argumente (vgl. Abschn. 1) ihre Kontur.

Natürlich kann aus den hier vorgebrachten (oder gleichsinnigen) Einwänden weder eine Bestreitung des theoretischen oder auch praktischen Nutzens des Programm-Paradigmas überhaupt, noch, weniger stark, eine solche seiner Fähigkeit zur Abbildung gewisser subjektiver, auch reflexiver, Strukturen abgeleitet werden - man braucht dazu nur auf Boden (1977, 1981) hinzuweisen. Ferner wurde ja in Abschn. 4.1 gezeigt, daß eine algorithmische Theorie des Beweisens im Feld der Aussagenlogik möglich ist; schließlich gehen durch das Prinzip der induktiven Definition reflexive Momente in die "rekursiven Funktionen" als einer möglichen Präzisierung des Algorithmuskonzepts ein (Hermes 1978, Kap. 5). Demgegenüber liegt ihr positiver Gehalt, sofern sie sich als widerlegungsresistent erweisen, darin, daß sie hinsichtlich der Phänomene 'Reflexivität' und 'Selbstbewußtsein' auf grundlegende strukturelle Unterschiede zwischen Menschen und programmgesteuerten Maschinen verweisen. Wenn also McCarthy schreibt:

"To ascribe certain 'beliefs', 'knowledge', ..., 'consciousness', 'abilities' ... to a machine or computer program is legitimate, when such an ascription expresses the same information about the machine that it expresses about a person. It is useful, when the ascription helps us understand the structure of the machine, its past or future behaviour, or how to repair or improve it" (McCarthy 1979, p. 161)⁸⁾,

so ist dagegen einzuwenden: Auch wenn es nützlich sein sollte, bei programmgesteuerten Maschinen von Reflexivität und Selbstbewußtsein zu reden (pragmatischer Aspekt), so wären solche Zuschreibungen nur von einer 'als ob'-Qualität; wegen der aufgezeigten Strukturdifferenzen käme ihnen nicht derselbe Gehalt wie im Humanbereich zu (ontologischer Aspekt).

Anmerkungen

1) Für eine kritische Lektüre einer ersten Fassung des Textes, Anmerkungen und Anregungen danke ich Dietrich Albert, Carl F. Graumann, Alexandre Métraux, Hans-Werner Wahl (Heidelberg) und Hans Lenk (Karlsruhe).

2) Insbesondere im Hinblick auf die Überlegungen zur Selbstreflexion beim Problemlösen (Abschn. 4.1) ist es zweckmäßig, auf die häufig gemachte Gegenüberstellung zwischen "Algorithmen" und "Heurismen" hinzuweisen (vgl. bspw. Hilgard & Bower 1966, p. 387 ff.; Boden 1977, p. 346/352; Dörner 1976, p. 37/49). Sie wird im Zusammenhang mit Problemlöseansätzen auf der Objektebene getroffen und bezieht sich auf die Differenz zwischen solchen Prozeduren, mit denen man geradewegs auf die letztendlich auch erreichte Lösung losgeht (Algorithmen, bspw. beim Lösen einer quadratischen Gleichung), und solchen, bei denen man gewisse allgemeine Strategien anwendet, die nicht so eng 'lösungsfixiert' sind, bspw. zur Problem-

transformation, wobei das transformierte Problem dann evtl. gelöst werden kann, oder zur Vermeidung der "brute-force"-Methode, bei der systematisch Schritt für Schritt alle Möglichkeiten durchgegangen werden (Heuristiken, bspw. die Rückführung auf schon bekannte Theoreme beim Beweisen oder die "breadth-first"- statt "depth-first"-Methode (Boden 1977, p. 348 ff.)). Diese Unterscheidung interferiert jedoch nicht mit dem Algorithmus-Konzept als theoretischem Kernstück der kognitiven Psychologie i.o.S., das zur Metaebene gehört. Konkret: eine voll explizierte Theorie des Problemlösens im Rahmen dieses Paradigmas wäre angebar als Programm S, in dem alle Übergänge von einem Zeitpunkt t zum nachfolgenden t + 1 durch den Input I zur Zeit t und S vollständig spezifiziert wären (vgl. Lachman et al. 1979, p. 104). Dies schließt nicht aus, daß innerhalb der Lösungsversuche auch Heuristiken zum Tragen kommen, nur wäre deren Auftreten und Ablauf voll und ganz spezifiziert (zur Relativität der Unterscheidung zwischen Algorithmen und Heuristiken auf Objektebene vgl. Boden 1977, p. 347 f., p. 351 f.).

3) Der Behauptung, daß Bewußtsein unzureichend gefaßt ist, wenn es nur objektbezogen als 'Bewußtsein-von' unter Abblendung seiner Subjekthaftigkeit als 'Bewußtsein-für' begriffen wird, wird in der gegenwärtigen Bewußtseinsdiskussion eine beträchtliche Bedeutung zugemessen. So gründet Nagel seine Kritik an materialistischen Reduktionsansätzen auf die nicht wegskamotierbare Subjektivität der Erfahrung (Nagel 1981, p. 262, p. 268 f.) und verlangt Dennett im Anschluß an Nagel ein funktionales Bewußtseinskonzept, das 'das Ich' angemessen repräsentiere (Dennett 1979c, p.149 f., p.153 f.).

4) Eine entsprechende Irreduzibilitätsbehauptung stammt von H.-N. Castañeda. Er beweist rein logisch-sprachanalytisch die Unmöglichkeit, das Erste-Person-Pronomen 'ich' in gewissen Sachverhaltssätzen salva propositione durch eine Außenbeschreibung in der dritten Person zu ersetzen, indem er zeigt, daß der Wahrheitswert des ich-Satzes nicht substitutionsinvariant ist (hier nach Castañeda 1968). Betrachten wir dazu bspw. den fiktiven "Editor of Soul", der (zum Zeitpunkt t) in oratio recta von sich behauptet: (1) "I am a millionaire". Dieser Sachverhalt läßt sich äquivalent wie folgt ausdrücken: (2) "The Editor of Soul believes (at time t) that he (himself) is a millionaire" (a.a.O., p. 261). Offensichtlich gilt, daß die Ersetzbarkeit von "ich" in (1) die Ersetzbarkeit des "er (selbst)" in (2) zur Folge hätte, denn beide Male ist ja der gemeint, der als "Herausgeber der Zeitschrift Soul" angesprochen ist. Um die o.a. Behauptung zu beweisen, reicht es also, die Nicht-Ersetzbarkeit salva propositione des sog. "Quasi-Indikators" "er (selbst)" (a.a.O., p. 262) zu zeigen. Dazu werden die verschiedenen Ersetzungsmöglichkeiten durchgegangen (bspw. durch eine definite Beschreibung oder einen Namen des Herausgebers), wobei sich stets Sachverhalts-Veränderungen auf-

weisen lassen (a.a.O., p. 262 f.). Nach Castañeda gelten diese Analysen auch unabhängig von dem angegebenen Beispiel (1): "Thus, more generally, no first-person proposition is identical with any third-person proposition, if both are in oratio recta" (a.a.O., p. 263).

5) Vor diesem Hintergrund ist gegen McCarthy einzuwenden, daß er bei seinem Konstruktionsversuch von "self-consciousness" fälschlich den Term "I" wie einen Namen zur Objektbezeichnung verwendet (McCarthy 1979, p. 183).

Dementsprechend ist in Winograds SHRDLU-Klötzchenwelt-Programm "I" schlicht ein weiterer Name, der äquivalent etwa durch SHRDLU ersetzt werden kann (Baker 1981, p. 161, FN 14).

6) Die vorgetragenen Überlegungen erlauben zwei weitere Folgerungen. Nach Baker (1981) ist es unangemessen, im Zusammenhang mit Computern von "Handeln" zu sprechen, denn: "Handeln" heißt, "Intentionen" formulieren zu können (a.a.O., p. 157). Dazu aber bedürfe es notwendig einer "irreduziblen" "EPP" (a.a.O., p. 157, p. 162 f.). Da die EPP Maschinen nicht zukomme (a.a.O., p. 157, 159, 162), vermöchten sie auch nicht zu handeln (a.a.O., p. 157). - Auf einen anderen Aspekt verweisen die Analysen von Schmitz. Danach bringen affektbezogene Ausdrücke in ich-Sätzen als besondere Nuance mein Betroffensein, mein Berührtsein als meines zum Ausdruck, das nicht auf eine Außenbeschreibung reduzierbar ist. Ist Maschinen/Programmen eine genuine EPP nicht zuzuschreiben, so können in bezug auf sie diese Sachverhalte nicht sinngetreu behauptet werden: Hinsichtlich ihrer speziellen Ich-Bezogenheit sind also Gefühle oder Affekte nicht qua Programm simulierbar.

7) Auch Frey (insbes. 1980, p. 61/65) bezieht sich auf realisierte Systeme, verwendet das Gödel-Argument aber auch in diesem Falle. Der Grundgedanke ist dabei der folgende (a.a.O., p. 70): Den sprachlichen Ausdrücken und Funktionen werden entsprechende Gehirnfunktionen "eindeutig" zugeordnet. Ein kybernetisches Modell des Bewußtseins müßte nun diese Gehirnfunktionen abbilden, so daß also ein Modell zumindest der Komplexität der Sprache möglich sein müßte - und dagegen kann das Gödel-Argument eingewandt werden (vgl. Abschn. 4.2). Das aber übersieht m.E. gerade das durch lernende Systeme gestellte Problem: Diese verfügen ja ab origine noch nicht über ein voll realisiertes Modell ihres Verhaltensbereiches, sondern entwickeln dies gerade erst in der Interaktion mit einer bereichsrelevanten Umwelt.

M.E. vermag auch Baker (1981) keinen Unmöglichkeit**s**beweis gegen die Möglichkeit selbstbewußter lernender Systeme zu führen. Sie nimmt die Hardware-Perspektive zwar auf (a.a.O., p. 161), stellt aber dann nur apodiktisch hin, daß Computern "Referenzfehler" unterlaufen können, gegen die

eigentlich selbstbewußte Wesen immun sind, bspw. kann niemand "ich" sagen und einen anderen meinen (a.a.O., p. 161), begründet aber nicht, warum dies auch bei lernenden Automaten notwendig so sein muß, sondern expliziert m.E. nur den Aspekt der EPP, daß "ich" kein Eigenname ist.

8) Zugrunde liegt Dennetts Konzeption "intentionaler Systeme" (Dennett 1979a, b).

Literaturverzeichnis

- Baker, L.R. Why computers can't act. American Philosophical Quarterly, 1981, 18, 157/163.
- Boden, M.A. Artificial intelligence and natural man. New York: Basic Books 1977.
- Minds and mechanisms. Philosophical psychology and computational models. Brighton, Sussex: Harvester 1981.
- Intentionality and physical systems. In: Boden 1981, p. 52/70 (orig. 1970) (a).
- The computational metaphor in psychology. In: Boden 1981, p. 31/51 (orig. 1979) (b).
- Brentano, F. Psychologie vom empirischen Standpunkt. Erstes Buch. Hrsgg. v. O. Kraus. Hamburg: Meiner 1955 (orig. 1874).
- Castañeda, H.-N. On the phenomeno-logic of the I. Akten des XIVten Internationalen Kongresses für Philosophie. Bd. 3. Wien: Herder 1968, 260/266.
- Dennett, D.C. Brainstorms. Philosophical essays on mind and psychology. Hassocks, Sussex: Harvester 1979.
- Intentional systems. In: Dennett 1979, p. 3/22 (orig. 1971) (a).
- Conditions of personhood. In: Dennett 1979, p. 267/85 (orig. 1976) (b).
- Toward a cognitive theory of consciousness. In: Dennett 1979, p. 149/173 (orig. 1978) (c).
- The abilities of men and machines. In: Dennett 1979, p. 156/166 (d).
- Dörner, D. Problemlösen als Informationsverarbeitung. Stuttgart: Kohlhammer 1976.
- Self-reflection and problem solving. In: F. Klix (Hrsg.). Human and artificial intelligence. Amsterdam: North-Holland Publ. Comp. 1979; p. 101/107.

- Dreyfus, H.L. What computers can't do: A critique of artificial reason. New York: Harper & Row 1972.
- Eisler, R. Wörterbuch der philosophischen Begriffe. Berlin: Mittler 1927 (4. A.; 1. A. 1900).
- Enzyklopädie Philosophie und Wissenschaftstheorie. Bd. 1. Hrsgg. von J. Mittelstraß et al. Mannheim: Bibliographisches Institut 1980.
- Feger, H. & Graumann, C.F. Beobachtung und Beschreibung von Erleben und Verhalten. In: Enzyklopädie der Psychologie. Serie I: Forschungsmethoden der Psychologie. Bd. 2. Datenerhebung. Hrsgg. von H. Feger & J. Bredenkamp. Göttingen: Hogrefe 1983, p. 76/134.
- Frey, G. Sprache - Ausdruck des Bewußtseins. Stuttgart: Kohlhammer 1965.
- Sind bewußtseinsanalogue Maschinen möglich? In: H.-W. Klement (Hrsg.) Bewußtsein. Ein Zentralproblem der Wissenschaften. Baden-Baden: Agis 1975.
- Theorie des Bewußtseins. Freiburg: Alber 1980.
- Gallup, G.G.jr. Self-recognition in primates. A comparative approach to the bidirectional properties of consciousness. American Psychologist 1977, 32, 329/338.
- Habermas, J. Erkenntnis und Interesse. (Mit einem neuen Nachwort). Frankfurt/M.: Suhrkamp 1973 (2.A.; 1.A. 1968).
- Hermes, H. Einführung in die mathematische Logik. Klassische Prädikatenlogik. Stuttgart: Teubner 1963.
- Aufzählbarkeit, Berechenbarkeit, Entscheidbarkeit. Einführung in die Theorie der rekursiven Funktionen. Berlin: Springer 1978 (3.A.; 1.A. 1961).
- Hilgard, E.R. & Bower, G.H. Theories of learning. New York: Appleton-Century-Crofts 1966 (3.A.; 1.A. 1948).
- Hofstadter, D.R. Gödel, Escher, Bach: An eternal golden braid. Harmondsworth: Penguin Books 1980.
- Husserl, E. Ideen zu einer reinen Phänomenologie und Phänomenologischen Psychologie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie. Hrsgg. v. W. Biemel. Husserliana Bd. 3. Haag: Nijhoff 1950 (orig. 1913).
- Kafka, F. Ein Bericht für eine Akademie. In: Sämtliche Erzählungen. Frankfurt/M.: Fischer 1969 (orig. 1917).
- Kutschera, F. von & Breitkopf, A. Einführung in die moderne Logik. Freiburg: Alber 1971.

- Lachman, R., Lachman, J.L. & Butterfield, E.C. Cognitive psychology and information processing. An introduction. Hillsdale, N.J.: Lawrence Erlbaum 1979.
- Lewis, M. & Brooks-Gunn, J. Social cognition and the acquisition of self. New York: Plenum Press 1979.
- Lorenz, K. Die Rückseite des Spiegels. Versuch einer Naturgeschichte menschlichen Erkennens. München: Piper 1973.
- Lucas, J.R. Minds, machines, and Gödel. Philosophy 1961, 36, 112/127.
- Malson, L., Itard, J. & Mannoni, O. Die wilden Kinder. Frankfurt/M.: Suhrkamp 1972.
- McCarthy, J. Ascribing mental qualities to machines. In: J. Ringle (ed.) Philosophical perspectives in Artificial Intelligence. Brighton, Sussex: Harvester 1979, p. 161/195.
- Mead, G.H. Mind, self, and society. Chicago: University of Chicago Press 1934.
- Nagel, E. & Newman, J.R. Gödel's proof. New York: New York University Press 1958.
- Nagel, Th. Wie ist es, eine Fledermaus zu sein? In: P. Bieri (Hrsg.) Analytische Philosophie des Geistes. Meisenheim/Glan: Hain 1981 (orig. 1974)
- Neisser, U. The imitation of man by machine. Science 1963, 139, 193/197.
----- Kognitive Psychologie. Stuttgart: Klett 1974 (orig. 1967).
- Newell, A. Physical symbol systems. In: Norman 1981, p. 37/85.
- Newell, A. & Simon, H.A. Human problem solving. Englewood Cliffs, N.J.: Prentice Hall 1972.
- Norman, D.A. (ed.) Perspectives on Cognitive Science. Hillsdale, N.J.: Erlbaum 1981.
----- Twelve issues for Cognitive Science. In: Norman 1981, p. 265/295 (a).
- Popper, K.R. & Eccles, J.C. Das Ich und sein Gehirn. München: 1982 (orig. 1977).
- Ristau, C.A. & Robbins, D. Language in the Great Apes. A critical review. In: J.S. Rosenblatt, R.A. Hinde, C. Beer & M.-C. Busnel (eds.) Advances in the study of behavior. Vol. 12. New York: Academic Press 1982.
- Schmitz, H. System der Philosophie. Dritter Band: Der Raum. Zweiter Teil: Der Gefühlsraum. Bonn: Bouvier 1969.

- Shelley, M.W. Frankenstein oder Der moderne Prometheus. Zürich: Manesse 1983 (orig. 1818).
- Stegmüller, W. Unvollständigkeit und Unentscheidbarkeit. Die metamathematischen Resultate von Gödel, Church, Kleene, Rosser und ihre erkenntnistheoretische Bedeutung. Wien: Springer 1959.
- Suppes, P. & Warren, H. On the generation and classification of defence mechanisms. International Journal of Psycho-Analysis 1975, 56, 405/414
- Tugendhat, E. Selbstbewußtsein und Selbstbestimmung. Sprachanalytische Interpretationen. Frankfurt/M.: Suhrkamp 1979.
- Wagner, H. Philosophie und Reflexion. München: Reinhardt 1959.
- Weizenbaum, J. Die Macht der Computer und die Ohnmacht der Vernunft. Frankfurt/M.: Suhrkamp 1982 (3.A.; orig. 1976).
- Wimmer, H. & Perner, J. Kognitionspsychologie. Eine Einführung. Stuttgart: Kohlhammer 1979.

Bisher erschienene Berichte aus dem

Psychologischen Institut der Universität Heidelberg

- Diskussionspapier Nr. 1: GROEBEN, N.: Vom behavioralen zum epistemologischen Subjektmodell: Paradigmawechsel in der Psychologie? September 1975
- Diskussionspapier Nr. 2: MÖBUS, C. & SIMONS, H.: Zur Fairness psychologischer Intelligenztests gegenüber ethnischen und sozialen Gruppen: Kritik klassischer Konzepte. Oktober 1975
- Diskussionspapier Nr. 3: WOTTAWA, H.: Skalenprobleme bei probabilistischen Meßmodellen. März 1976
- Diskussionspapier Nr. 4: TREIBER, B. & PETERMANN, F.: Zur Interaktion von Lernermerkmalen und Lehrmethoden: Rekonstruktion und Normierung des ATI-Forschungsprogramms. April 1976
- Diskussionspapier Nr. 5: MÖBUS, C. & WALLASCH, R.: Zur Erfassung von Hirnschädigungen bei Kindern: Nichtlineare Entscheidungsregeln auf der Basis von Veränderungsmessungen. August 1976
- Diskussionspapier Nr. 6: SCHEELE, B. & GROEBEN, N.: Voraussetzungs- und zielspezifische Anwendung von Konditionierungs- vs. kognitiven Lerntheorien in der klinischen Praxis. Dezember 1976
- Diskussionspapier Nr. 7: MÖBUS, C.: Zur Analyse nichtsymmetrischer Ähnlichkeitsurteile: Ein dimensionales Driftmodell, eine Vergleichshypothese, TVERSKY's Kontrastmodell und seine Fokushypothese. Juni 1977
- Diskussionspapier Nr. 8: SIMONS, H. & MÖBUS, C.: Veränderung von Berufschancen durch Intelligenztraining. Juli 1977
- Diskussionspapier Nr. 9: BRAUNMOHL, C. v. & GRIMM, H.: Zur Kommunikationspsychologie: Über Versuche der methodischen Konstitution eines genuin humanwissenschaftlichen Forschungsansatzes zur Entwicklung der Verständigungsfähigkeit. November 1977
- Diskussionspapier Nr. 10: HOFER, M.: Entwurf einer Heuristik für eine theoretisch geleitete Lehrer- und Erzieherbildung. November 1977
- Diskussionspapier Nr. 11: SCHEIBLER, D. & SCHNEIDER, W.: Probleme und Ergebnisse bei der Evaluation von Clusteranalyse-Verfahren. Juni 1978
- Diskussionspapier Nr. 12: SCHEELE, B.: Kognitions- und sprachpsychologische Aspekte der Arzt-Patient-Kommunikation. September 1978

Diskussionspapier Nr. 13: TREIBER, B. & SCHNEIDER, W.: Mehr-
ebenenanalyse sozialstruktureller Bedingungen schulischen
Lernens. Oktober 1978

Diskussionspapier Nr. 14: AHRENS, H.-J. & KORDY, H.: Möglich-
keiten und Grenzen der theoretischen Aussagekraft von multi-
dimensionalen Skalierungen bei der Untersuchung menschlicher
Informationsverarbeitung. Teil I: Formale und wissenschafts-
theoretische Grundlagen. März 1979

Diskussionspapier Nr. 15: GROEBEN, N.: Entwurf eines Utopie-
prinzips zur Generierung Psychologischer Konstrukte. Juni 1979

Diskussionspapier Nr. 16: WEINERT, F.E. & TREIBER, B.: School
Socialization and cognitive development. Juni 1979

Diskussionspapier Nr. 17: GUNDLACH H.: Inventarium der älteren
Experimentalapparate im Psychologischen Institut Heidelberg
sowie einige historische Bemerkungen. 1978

Diskussionspapier Nr. 18: SCHEELE, B. & GROEBEN, N.: Zur Rekon-
struktion von subjektiven Theorien mittlerer Reichweite.
Eine Methodik-Kombination von halbstandardisiertem Interview
(einschließlich Konfrontationstechnik) und Dialog-Konsens über
die Theorie-Rekonstruktion mittels der Struktur-Lege-Technik
(SLT). Dezember 1979

Diskussionspapier Nr. 19: GLOGER-TIPPELT, G.: Subjektive Theorie
von Frauen über ihre erste Schwangerschaft: Theoretische Kon-
zepte und methodische Möglichkeiten. Januar 1980

Diskussionspapier Nr. 20: KÄMMERER, A.: Das Konzept 'psycho-
therapeutische Strategie' am Beispiel des Problemlösens.
Juli 1980

Diskussionspapier Nr. 21: SCHEELE, B.: (unter Mitarbeit von
B.Tuschen und C.Maier): Subjektive Theorien über Ironie - als
Heuristik für einen wissenschaftlichen Hypothesenkörper.
August 1980

Diskussionspapier Nr. 22: TREIBER, B.: Erklärung von Förderungs-
effekten in Schulklassen durch Merkmale subjektiver Unter-
richtstheorien ihrer Lehrer. Oktober 1980

Diskussionspapier Nr. 23: RÖHRLE, B. & KOMMER, D.: Handlungs-
theoretische Betrachtungen zur primären Prävention psychi-
scher Störungen. Februar 1981

Diskussionspapier Nr. 24: VOIGT, F.: Die Entwicklung des Zahl-
begriffs. Teil I: Entwicklungslinien des Zahlbegriffs im
Vorschulalter: Übersicht über theoretische Probleme und empi-
rische Untersuchungen, mit einer Bibliographie zur Zahlbe-
griffsentwicklung. Teil II: Entwicklungslinien des Zahlbe-
griffs im Vorschulalter: Deskriptive Untersuchung des kind-
lichen Zahlverständnisses und verwandter Konzepte. April 1981

- Diskussionspapier Nr. 25: SCHNEIDER, G. & WEIMER, E.: Aspekte der Kategorisierung städtischer Umwelt - Eine empirische Untersuchung. Juni 1981
- Diskussionspapier Nr. 26: SCHNEIDER, W. & SCHEIBLER, D.: Zur Evaluation numerischer Klassifikation: Probleme beim Vergleich von Clusteranalysen. August 1981
- Diskussionspapier Nr. 27: DRINKMANN, A. & GROEBEN, N.: Techniken der Textorganisation zur Verbesserung des Lernens aus Texten: Ein metaanalytischer Überblick. November 1981
- Diskussionspapier Nr. 28: GRAUMANN, C.F.: Theorie und Geschichte. November 1982, Historische Reihe Nr. 1
- Diskussionspapier Nr. 29: WOODWARD, W.R.: From the Science of Language to *Völkerpsychologie*: Lotze, Steinthal, Lazarus, and Wundt. November 1982, Historische Reihe Nr. 2
- Diskussionspapier Nr. 30: SOMMER, J.: Dialogische Forschungsmethoden. Dezember 1982
- Diskussionspapier Nr. 31: WINTERMANTEL, M. & CHRISTMANN, U.: Textverarbeitung: Empirische Untersuchung zum Verstehen einer Personbeschreibung. Januar 1983
- Diskussionspapier Nr. 32: SCHMALHOFER, F.: Text Processing with and without Prior Knowledge: Knowledge- versus Heuristic-Dependent Representations. Februar 1983
- Diskussionspapier Nr. 33: METRAUX, A.: Victor de l'Aveyron oder Zum Streit zwischen Kulturalisten und Biologen am Anfang des 19. Jahrhunderts. Mai 1983, Historische Reihe Nr. 3
- Diskussionspapier Nr. 34: GRAUMANN, C.F.: Wundt - Bühler - Mead - Zur Sozialität und Sprachlichkeit menschlichen Handelns. Mai 1983, Historische Reihe Nr. 4
- Diskussionspapier Nr. 35: GUNDLACH, H.: Folk Psychology and Social Psychology oder? Das Los des Ausdrucks 'Völkerpsychologie' in den englischen Übersetzungen der Werke Wundts. Mai 1983, Historische Reihe Nr. 5
- Diskussionspapier Nr. 36: WOODWARD, W.R.: Hermann Lotze's Concept of Function: Its Kantian Origin and its Impact on Evolutionism in the United States. Mai 1983, Historische Reihe Nr. 6