

**BERICHT  
aus dem  
PSYCHOLOGISCHEN INSTITUT  
DER UNIVERSITÄT HEIDELBERG**

Stefan E. Hormuth

Methoden für psychologische Forschung im Feld:  
*Erfahrungsstichprobe, Autophotographie und Telefoninterview*

Februar 1985

Diskussionspapier Nr. 43

Stefan E. Hormuth

Methoden für psychologische Forschung im Feld:  
Erfahrungsstichprobe, Autophotographie und Telefoninterview

Februar 1985

Diskussionspapier Nr. 43

Psychologisches Institut der Universität Heidelberg

Inhalt

Hormuth, S. E. Experiment und Quasi-Experiment in der Sozialpsychologie: Die Notwendigkeit eines multiplen Methodenansatzes.

Hormuth, S.E. The random sampling of experiences in situ.

Hormuth, S.E. & Lalli, M. The role of urban environments for the self-concept: A research example using an autographical approach.

Hormuth, S. E. & Brückner, E. Telefoninterviews in Sozialforschung und Sozialpsychologie: Ausgewählte Probleme der Stichprobengewinnung, Kontaktierung und Versuchsplanung.

Dieses Diskussionspapier enthält, zum ersten Mal unter den Berichten aus dem Psychologischen Institut der Universität Heidelberg, verschiedene Beiträge. Ihnen allen ist gemeinsam, daß sie Erfahrungen mit verschiedenen Methoden der Feldforschung in der Sozialpsychologie berichtet. Sie sind hier in einem Bericht zusammengefaßt aus der Überzeugung heraus, die im ersten Beitrag ausführlicher begründet wird, daß es für die psychologische Forschung im Feld keine einzig angemessene Methode gibt. Vielmehr steht dem Forscher heute ein Repertoire von Methoden zur Verfügung, je mit den ihnen eigenen Stärken und Schwächen, die je in unterschiedlicher, der Fragestellung angepaßter Kombination eingesetzt werden können. Um nicht den Eindruck zu erwecken, daß eine einzelne, für sich hier dargestellte Methode als die vom Verfasser allein bevorzugt betrachtet wird, wurden diese Aufsätze in dieser Form zusammengefaßt.

Die hier vorgestellten Methoden werden im Projekt "Ortswechsel als Gelegenheit zur Änderung des Selbstkonzepts" (DFG Az.: Ho 905/1-1f), das im Rahmen des DFG-Schwerpunktprogrammes "Psychologische Ökologie" gefördert wird, verwendet. Für andere Fragestellungen würden sich vielleicht andere Methoden anbieten. Zugleich sind aber auch die hier behandelten in anderen Bereichen der Psychologie erfolgreich eingesetzt worden: so z. B. die Erfahrungsstichprobe in der Persönlichkeitspsychologie, die Autophotographie in der phänomenologischen Psychologie, die Telefonumfrage in der kognitiven Sozialpsychologie.

Die einzelnen Beiträge dieses Berichts können alle noch in fälligen Überarbeitungen aus Kritik und Rückmeldung profitieren. Insbesondere trifft dies für die zwei kürzeren Beiträge zu, die auf Vortragsmanuskripten beruhen. Der Verfasser ist deswegen für kritische Kommentare sehr dankbar.

Experiment und Quasi-Experiment in der Sozialpsychologie:

Die Notwendigkeit eines multiplen Methodenansatzes

Stefan E. Hormuth

Universität Heidelberg

Vortrag auf dem 1. Hamburger Symposium zur Methodologie der  
Sozialpsychologie, 18./19. Januar 1985. Adresse des Autors: Dr. Stefan  
Hormuth, Psychologisches Institut, Universität Heidelberg, Hauptstr. 47-  
51, D-6900 Heidelberg 1.

### Experiment und Quasi-Experiment in der Sozialpsychologie

Die Sozialpsychologie verstand sich in den sechziger Jahren als eine ausschließlich experimentelle Disziplin. Nach einer Phase, die häufig als die "Krise" der Sozialpsychologie bezeichnet wurde und die als solche Mitte der siebziger Jahre durch ein Machtwort von Marcia Guttentag beendet wurde, wurden auch häufiger Arbeiten als Sozialpsychologie akzeptiert, die nicht-experimentell angelegt waren.

Das Experiment hat in der Entwicklung der wissenschaftlichen Psychologie und der Sozialpsychologie einen besonderen Rang eingenommen. Während für Wundt die experimentelle Methode definierend in der Begründung einer Wissenschaft des Bewußtseins war (Vorwerg, 1962), wurde dies in den Jahren zwischen Triplets erstem Experiment (1897) und Moedes "Experimenteller Massenpsychologie" (1920) (ursprünglich war der Titel "Experimentelle Sozialpsychologie" vorgesehen aber wohl auf Anregung des Verlegers geändert worden) bzw. Floyd Allports "Social Psychology" (1924) auch ein wesentliches Element in der Begründung der Sozialpsychologie, die sich inzwischen auch deutlich als eine Psychologie des Individuums entwickelt hatte, auch wenn nicht mehr das Bewußtsein, sondern das Verhalten Gegenstand der wissenschaftlichen Untersuchung war.

Gerade dies wird dem Experiment heute zum Vorwurf gemacht, nämlich daß es nur die Wissenschaftlichkeit der Psychologie begründen helfen solle, aber zur Erforschung der eigentlichen Inhalte der Sozialpsychologie ungeeignet sei. Dieser Vorwurf ist möglicherweise im Bereich der Sozialpsychologie häufiger zu hören als in manchen anderen Bereichen der Psychologie.

Das sozialpsychologische Experiment, wie es sich heute darstellt, ist allerdings nur zum Teil dem Modell des Experiments in den Naturwissenschaften entlehnt. Experimente in den Naturwissenschaften

können unter solch kontrollierten, 'sterilen' Bedingungen stattfinden, wie dies für das sozialpsychologische Labor nie angenommen wurde. Deswegen wurden weitere Prinzipien des Experimentierens aus der landwirtschaftlichen Forschung übernommen, nämlich die Schaffung gleichartiger Gruppen durch Zufallszuweisung, und die darauf folgende Einführung entweder verschiedenartiger Treatments oder das Weglassen des Treatments in einer der Gruppen. Auch der faktorielle Versuchsplan ebenso wie die dazugehörigen statistischen Verfahren der Varianzanalyse entstammen der agrikulturellen Forschung, die mit der sozialpsychologischen das Anliegen bzw. die Notwendigkeit teilt, Wachstum bzw. Verhalten nicht unter völlig sterilen Bedingungen zu untersuchen.

Eine Untersuchung der These, daß das Experiment die Entwicklung der Sozialpsychologie zurückgehalten habe, muß davon ausgehen, daß die Methoden einer Wissenschaft und die Inhalte, mit denen sie sich befaßt, eng miteinander verknüpft sind. Dieser Ausgangspunkt ist sicher unbestritten. Die Kritik setzt da ein, wo das Festhalten an einer bestimmten Methode, hier dem Experiment, die Möglichkeiten der inhaltlichen und theoretischen Weiterentwicklung hemmen würde.

Allerdings muß dazu bemerkt werden, daß die Phase der ausschließlichen Konzentration auf das Experiment in der Sozialpsychologie außerordentlich kurz war. Bis in die sechziger Jahre hinein waren Felduntersuchungen, Beobachtungen, teilnehmende Beobachtungen und andere Methoden selbstverständlicher Bestandteil des sozialpsychologischen Methodenrepertoires. Auf der anderen Seite war das Experiment für diejenigen, die ganzheitliche und Feldforschungsmethoden durch ihr Verständnis der Determinanten des Verhaltens, z.B. feldtheoretisch, ebenso wie durch Anspruch auf Anwendbarkeit sozialpsychologischer Erkenntnisse begründeten ebenfalls selbstverständlicher Bestandteil ihres Methodenrepertoires, allen voran

Lewin. Die Dominanz des Experiments als Methode der Sozialpsychologie begann etwa mit der Dissonanztheorie (Festinger, 1957), wobei die Blützeit vielleicht von der Kritik von Chapanis und Chapanis (1964) an den Experimenten der Dissonanztheorie und den darauffolgenden methodischen Verbesserungen angesetzt werden kann. Das sozialpsychologische Experiment ist durch Aronsons und Carlsmiths (1969) Artikel im Handbook of Social Psychology am besten gekennzeichnet. Aber 1969 hatte bereits die Kritik an der Methode ebenso wie an den Inhalten voll eingesetzt, zum Beispiel im Austausch von Ring (1967) und McGuire (1967). Auf dem Weltkongreß 1972 in Tokio hielt McGuire in seinem Yin und Yang-Vortrag (1973 veröffentlicht) den Abgesang auf die Ära der Dominanz des Laborexperimentes in der Sozialpsychologie, in den andere, z.B. Gergen (1973, 1978) mit einstimmten.

Ebenfalls 1972 wurde auch von europäischer Seite der Kontext der Sozialpsychologie in einem gleichnamigen Buch, herausgegeben von Israel und Tajfel, kritisch beurteilt. Tajfel setzte sich in diesem Band kritisch mit dem Experiment auseinander, allerdings auf der Grundlage einer Voraussetzung: "...there is no evidence that other approaches to psychological aspects of social conduct present even as much solidity as the experimental straw appears to have" (p. 69). Man kann die Anfangs- und Endpunkte des Zeitraumes, in dem die experimentelle Laborforschung die via regia sozialpsychologischer Methodik war, vielleicht ein wenig verschieben: Ich komme auf fünf Jahre, eine wahrhaft kurze Zeitspanne auch in der kurzen Geschichte der Sozialpsychologie.

Die Diskussion über die Bedeutung des Experiments in der Sozialpsychologie liesse sich hier und heute leicht weiterführen, mit bekannten Argumenten, ein wenig neuartig klingenden dazu, so wie es häufig noch getan wird, wenn Sozialpsychologen zusammen kommen. Meiner Meinung nach ist die Diskussion in dieser Form ausgestanden und hat zu der Akzeptanz neuer Inhalte und Methoden geführt. Die Diskussion über

die Sozialpsychologie als rein experimentelle Disziplin wurde in den späten sechziger, frühen siebziger Jahren geführt, und hat seither etwas an Originalität verloren.

Die Sozialpsychologie hat sich dadurch verändert, auch wenn es einfach bei der Lektüre ihrer hauptsächlichen Zeitschriften nicht immer so ausschauen mag. Hier ist nach wie vor die experimentelle Sozialpsychologie an Anzahl der Veröffentlichungen dominierend. Gleichwohl stehen Felduntersuchungen gleichberechtigt daneben. Darüber hinaus hat sich die Sozialpsychologie eine Reihe neuer Organe geschaffen, z.B. das Journal of Basic and Applied Psychology, Journal of Social and Clinical Psychology, Social Psychology and Language, Journal of Social and Personal Relationships, in denen sozialpsychologische Arbeiten veröffentlicht werden, die vom Inhalt und den Methoden her die Vielfalt der Sozialpsychologie viel eher aufzeigen, als dies die traditionell führenden Zeitschriften allein tun können. Manche Psychologen, von der Ausbildung her Sozialpsychologen, sind abgewandert in andere Bereiche, zum Beispiel in die Organisationspsychologie, die Gemeindepsychologie, die Umweltpsychologie, um nur einige zu nennen, wo sie methodisch und inhaltlich vielfältigere Arbeiten leisten.

Die Kritik am Laborexperiment kann man vielleicht nach drei Zielorientierungen einteilen: (1) Abschaffung des Experimentes, (2) Verbesserung des Laborexperimentes, und (3) Erweiterung der Möglichkeiten auf der Grundlage der experimentellen Methode.

Die erste Gruppe, gekennzeichnet zum Beispiel durch Forderungen nach erhöhter sozialer Relevanz und der Verwirklichung von Aktionsforschung, ist häufig auch mit der Forderung nach qualitativen Forschungsmethoden identifiziert. Die zweite Zielorientierung, nämlich die Verbesserung des Laborexperimentes, begann schon in den frühen sechziger Jahren mit empirischer Forschung zur Sozialpsychologie des Experimentes, so zum Beispiel Ornes (1962) Untersuchungen zum

Aufforderungscharakter der experimentellen Situation, die bestimmte Hinweisreize für die Versuchsperson enthalten mag, oder Rosenthals (1966) Arbeiten zum Versuchsleitereffekt. Die Artefaktforschung leistet seither, und auch immer noch, einen Beitrag zum Verständnis der sozialen Situation im sozialpsychologischen Experiment, wodurch die interne Validität dieses Forschungsinstrumentes verbessert werden kann.

Die dritte Zielrichtung, nämlich die Erweiterung der Möglichkeiten des Experimentes, ging davon aus, daß die Logik des Experimentierens den besten Zugang zum Nachweis von Kausalität bietet, daß die dazu verwendete Kontrolle und Isolation allerdings den Zugang zu wesentlichen Fragestellungen verwehrt, nämlich solchen, die außerhalb des Labors liegen und dadurch einen wissenschaftlichen Zugang unter natürlichen Bedingungen verlangen, entweder durch experimentelle Eingriffe in diese natürlichen Bedingungen oder durch Nutzung der natürlich vorkommenden Variationen in den Bedingungen des Verhaltens. Feldforschung sollte unter kontrollierteren Bedingungen als bislang möglich sein. Mit diesem Ansatz verbindet sich vor allem der Name Don Campbells und seine Arbeiten zum Quasi-Experiment (Campbell & Stanley, 1966; Cook & Campbell, 1976; Cook & Campbell, 1979).

Lassen Sie mich zuerst noch einmal die Logik des Experiments, und damit auch des Feldexperiments, zusammenfassen. Der klassische randomisierte Versuchsplan mit einer Treatment-Gruppe und einer Kontrollgruppe, die beide einem Vortest und einem Nachtest unterworfen wurden, gilt allgemein als der beste experimentelle Versuchsplan zur Untersuchung kausaler Hypothesen. Sein Verdienst ist die Einfachheit, mit der Fragen über Kausalität dadurch beantwortet werden können. Die grundlegende Frage "Ist das Treatment mit dem intendierten Effekt verbunden?" kann durch Veränderungen in der Treatment-Gruppe zwischen Vor- und Nachtest beantwortet werden. Die schwierigere Frage "Könnte diese Veränderung hervorgerufen sein durch andere Faktoren als dem

Treatment?" kann zumindest teilweise durch einen Vergleich der Veränderungen zwischen Vortest und Nachtest in der treatment und Kontrollgruppe beantwortet werden. Wenn diese Veränderungen unterschiedlich sind, können alle Einflüsse ausgeschlossen werden, die sich gleichermaßen auf beide Gruppen auswirken sollten, zum Beispiel Selektion, Testeffekte, statistische Regression, Reifung, und ähnliche.

Die Logik des Quasi-Experiments unterscheidet sich von der des Experiments darin, daß nicht von einer randomisierten Zuteilung zu den experimentellen Bedingungen ausgegangen werden kann. Stattdessen beruhen quasi-experimentelle Designs auf verschiedenen Methoden, vor allem der Verwendung nicht-äquivalenter Kontrollgruppen und mehrfacher Messungen, um danach beurteilen zu können, ob Veränderungen stattgefunden haben, und ob diese Veränderungen den experimentell geschaffenen Bedingungen zuzuschreiben sind oder alternativ erklärt werden können.

Um eine sinnvolle kausale Analyse unter den Bedingungen des Fehlens von Zufallszuweisung durchführen zu können, sind allerdings, so Campbell, zwei weitere Bedingungen notwendig: Vortests, um Veränderungen über die Zeit feststellen zu können, und Kontrollgruppen, um die Veränderungen durch das experimentelle Treatment unterscheiden zu können von spontanen Veränderungen und solchen, die durch weitere externe Einflüsse aus der Ökologie der Untersuchungsteilnehmer auf diese einwirken.

Quasi-experimentelle Versuchspläne sind nicht dazu gemacht, experimentelle Versuchspläne zu ersetzen. Vielmehr sind sie dann einzusetzen, wenn entweder (1) ein als Experiment geplanter Versuchsplan nicht durchhaltbar ist, zum Beispiel aufgrund ungleichmäßigen, das heißt nicht zufälligen, Ausfalls der Untersuchungsteilnehmer. Hier können bei randomisierten (Feld)experimenten Probleme antizipiert werden, die dann aufgrund quasi-experimenteller Prinzipien bearbeitet werden können. (2)

Manchmal ist von vornherein eine zufällige Zuweisung zu treatment- und Kontrollbedingungen nicht möglich. Hier kann der Forscher versuchen, eine oder mehrerer Kontrollgruppen zu gewinnen, die der treatment-Gruppe in manchen Variablen auf bekannte Weise ähnlich oder unähnlich sind. Hier sind Versuchspläne mit sogenannten "nicht-vergleichbaren Kontrollgruppen" angezeigt, ebenso wie multiple Messungen oder direkte Messungen der Bedrohungen der Validität. Schließlich gibt es auch Forschungsprobleme, bei denen von vornherein die Untersuchung oder Bildung vergleichbarer oder nicht-vergleichbarer Kontrollgruppen nicht möglich ist. Hier sind dann zum Beispiel Zeitreihenanalysen angemessen (vgl. Hormuth, Fitzgerald, & Cook, 1985).

Das Quasi-Experiment ist den meisten wohl noch aus der ursprünglichen Veröffentlichung von Campbell und Stanley (1966) bekannt. Dort wurde eine recht klare Liste der Gefährdungen interner und externer Validität vorgestellt zusammen mit einer Reihe von Versuchsplänen, die Validität, vor allem aber interne Validität, unter suboptimalen Bedingungen zu verbessern helfen sollten.

Die Methodologie des Quasiexperiments, wie sie sich heute nach beinahe zwanzig Jahren darstellt, ist nicht mehr, wie bei Campbell und Stanley, eine Aufzählung von Gefährdungen der Validität und eine Reihe von Versuchsplänen, die fehlende Randomisierung auszugleichen versuchen. Die methodologische Erweiterung wird besonders deutlich in den Arbeiten mit und von Cook (Cook & Campbell, 1976, 1979; Cook, 19xx; Hormuth, Fitzgerald & Cook, 1985). Vielmehr handelt es sich inzwischen um ein System von Kriterien zur Beurteilung der Validität der Erfassung komplexer kausaler Beziehungen unter suboptimalen Forschungsbedingungen. In anderen Worten: Ausgehend von der Idealvorstellung des randomisierten Experiments umfaßt Quasi-Experimentieren die Antizipation von Fehlschlägen und Gefährdungen von Versuchsplänen und wie man diesen begegnen kann. Hierzu gehören Möglichkeiten der Einführung zusätzlicher

Variablen, der Einführung zusätzlicher Kontrollgruppen, ebenso wie zusätzlicher Analysen der Daten. Durch den Versuch, ein Problem auf verschiedene Weise anzugehen, sollen gegenseitige Validierungen auf verschiedene Weise die kausale Beweisführung einkreisen. Lassen Sie mich hierauf etwas ausführlicher eingehen.

Der Kausalitätsbegriff des Quasi-Experiments ist ein pragmatischer. Das Quasi-Experiment versucht nicht, kausale Abhängigkeiten vollständig zu beschreiben. Im Gegensatz zu manchen multivariaten Ansätzen, die komplexe Abhängigkeiten postulieren und diese z.B. in Pfad- oder LISREL-Modellen, darzustellen versuchen, geht das Quasi-Experiment davon aus, daß es sinnvoll ist, einige wenige Variablen zu isolieren, von denen anzunehmen ist, daß sie einen Unterschied machen, ohne deswegen die komplexen Abhängigkeiten vollständig aufzuklären zu wollen. Dies wird vor allem da wesentlich, wo der Forscher die empirisch gewonnenen Erkenntnisse in Interventionsmöglichkeiten umsetzen will. Insofern handelt es sich um einen pragmatischen Kausalitätsbegriff, der sicher nicht perfekt ist, meiner Meinung nach aber im Vergleich mit der geringen Stabilität komplexer kausaler Modelle (und deren Möglichkeiten der Anpassung des fit) durchaus bestehen kann. Selbstverständlich schließt die Anwendung quasi-experimenteller Prinzipien die Bildung und Analyse komplexer kausaler Modelle nicht aus.

Die Grundlage quasi-experimenteller Methodik liegt im Versuch der Falsifikation der Hypothesen aufgrund der Plausibilität alternativer Hypothesen. Die durch Campbell und Stanley aufgeführten Gefährdungen der internen und externen Validität von Versuchsplänen sind nichts als der Versuch, eine Reihe von plausiblen alternativen Hypothesen zur Verfügung zu stellen, die in den Sozialwissenschaften besonders häufig zu berücksichtigen sind. Keineswegs stellen sie den Anspruch auf Ausschließlichkeit oder können nicht auch durch andere plausible Alternativhypothesen ergänzt oder ersetzt werden, die aus den

theoretischen, inhaltlichen oder wissenschaftstheoretischen Annahmen des Forschers begründbar sind.

Die Methode des Quasi-Experiments ist weiterhin pragmatisch und nicht absolut in ihren Annahmen über die Meßbarkeit psychologischer und sozialer Tatbestände. Danach sind Messungen dann am zuverlässigsten, wenn sie aufgrund verschiedener Meßinstrumente und möglicherweise über verschiedene Beobachter zu ähnlichen Ergebnissen kommen. Jedes Meßinstrument für sich hat inhärente Schwächen, die nur durch Verwendung verschiedenartiger Meßinstrumente oder Variablen gegenseitig ausgeglichen werden können. Dies überträgt sich auch auf die Verwendung von Ansätzen der Datenanalyse, auf die Argumentation aufgrund einzelner Untersuchungen, oder sogar auf die Ergebnisse der Untersuchungen einzelner Forscher. Jede Konzentration auf einen einzelnen Ansatz ist nicht in der Lage, die spezifischen Schwächen dieses Ansatzes ohne Rückgriff auf andere Methoden auszugleichen. Die Logik des Quasi-Experiments ist damit zur Logik eines multiplen Methodenansatzes zur Forschung unter suboptimalen Bedingungen (wie sie in den Sozialwissenschaften vorherrschen) geworden.

Quasi-experimentelle Ansätze haben große Bedeutung außerhalb des engeren Gebietes der Sozialpsychologie gewonnen. Dies entspricht durchaus der Intention Campbells, der nicht einen methodischen Ansatz für die Sozialpsychologie, sondern für die Sozialwissenschaften allgemein erarbeiten wollte. Mit dem methodischen Ansatz verband er Vorstellungen über eine völlig neue gesellschaftliche Rolle sozialwissenschaftlicher Forschung auf der Grundlage des Experiments (Campbell, 1969) und seiner Variante unter suboptimalen Bedingungen, des Quasi-Experiments. Diese Vorstellungen Campbells und anderer haben sich, vor allem in den USA, teilweise verwirklicht, und zur Bildung der Evaluationsforschung als einer neuen Disziplin angewandter Sozialforschung geführt. Komplexe sozialwissenschaftliche Fragestellungen,

wie sie zum Beispiel in der Evaluationsforschung bearbeitet werden, haben darüber hinaus gezeigt, daß die dort üblichen Fragestellungen den multiplen Methodenansatz am ehesten in interdisziplinärer Zusammenarbeit verwirklichen können.

Sowohl die sozialpolitischen Anliegen der Evaluationsforschung als auch der dort notwendige und praktizierte Ansatz interdisziplinärer Zusammenarbeit gehen weit über die Möglichkeiten, Anliegen und Zielvorstellungen der eigentlichen Sozialpsychologie hinaus. Dies soll deswegen hier nicht als Modell angemessener sozialpsychologischer Methodik vorgestellt werden, vielmehr soll die Bedeutung der quasi-experimentellen Methodologie im engeren Rahmen unserer Disziplin "Sozialpsychologie" betrachtet werden unter Anerkennung der Tatsache, daß diese Disziplin sich zwar inhaltlich und methodisch in den letzten zehn, fünfzehn Jahren weiterentwickelt hat und weiterentwickeln mußte, daß es aber weiterhin sinnvoll ist, ihren Gegenstand zu beschränken auf das Verhalten und Erleben des Individuums in seiner sozialen Abhängigkeit, eine Definition, die meiner Meinung nach die Aufgaben der Sozialpsychologie noch auf lange Zeit hin angemessen beschreiben kann. Bei der Methode gehe ich davon aus, daß die Sozialpsychologie eine empirische Wissenschaft ist, die kausale Erklärungen sucht.

Inerhalb dieser Disziplin können wir versuchen, die Sichtweise der Quasi-Experimentalisten für die Durchführung und Bewertung empirischer Untersuchungen für uns nutzbar zu machen. Eine Beschreibung des gegenwärtigen Zustandes der Disziplin muß davon ausgehen, daß eine Vielzahl von quantitativen und qualitativen Methoden heute Anwendung finden. Das Experiment ist zwar nach wie vor die dominierende Methode (und das aus gutem Grund), vor allem im Bereich der sozialen Kognition, aber nicht die einzige. Daneben versuchen sich eine Vielzahl von Methoden zu etablieren, teilweise solche die in der Psychologie seit langem bekannt sind, wie das Interview, teilweise solche, die aus

Nachbarwissenschaften übernommen werden, wie die Umfrage, teilweise solche, die - auch aufgrund technologischer Entwicklungen - neu- oder weiterentwickelt wurden, wie z. B. die Erfahrungsstichprobe, die Zufallsstichproben von Alltagsverhalten *in situ* ermöglichen soll.

Ausgehend von den Erfahrungen des Quasi-Experimentes und der Annahme, daß sozialpsychologische Forschung häufig unter 'suboptimalen Bedingungen' durchgeführt werden muß, möchte ich einige Empfehlungen für die sinnvolle methodische Weiterentwicklung des Faches aussprechen:

- 1) Der Gegenstand der Sozialpsychologie kann nicht auf den Zugang durch eine einzelne Methode beschränkt werden.
- 2) Die verschiedenen methodischen Zugänge zum Gegenstand der Sozialpsychologie haben je unterschiedliche Stärken und Schwächen, die meist als unausgeglichenes Verhältnis zwischen interner und externer Validität beschrieben werden können.
- 3) In Erkenntnis dessen kann das Vertrauen in eine Aussage der Sozialpsychologie nicht auf dem Zugang durch eine einzelne Methode beruhen.
- 4) Die Stärken und Schwächen verschiedener Methoden sind ebenso empirisch zugänglich. Die empirische Erfassung der Methodenqualität kann und soll auch untersuchungsbegleitend geschehen, da die Qualität sich in verschiedenem Kontext unterschiedlich ausprägen kann.
- 5) Ein angemessener methodischer Zugang muß sich die unterschiedlichen Stärken verschiedener Methoden durch den Einsatz verschiedener Methoden zur Erforschung eines Problems oder einer Hypothese zunutze machen.

Das Vertrauen in eine Aussage der Sozialpsychologie kann immer nur probabilistisch sein. Die Bedeutung der einzelnen Untersuchung, des einzigartigen methodischen Zugangs relativiert sich dadurch. Es erhöht sich durch die Verwendung multipler Tests auf der Ebene der Variablen, der Methoden, der Kontexte, und auch der Forscher.

## Literatur

- Allport, F. H. (1924). Social Psychology. Boston: Houghton Mifflin
- Aronson, E., & Carlsmith, J. M. (1969). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology, 2nd ed., Vol. II, (pp. 1-79). Reading, MA: Addison-Wesley.
- Campbell, D. T. (1969). Reforms as experiments. American Psychologist, 24, 409-429.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook for research on teaching. Chicago: Rand McNally.
- Chapanis, N. & Chapanis, A. (1964). Cognitive dissonance: Five years later. Psychological Bulletin, 61, 1-22.
- Cook, T. D. (im Druck). Quasi-experimentation: Its ontology, epistemology, and methodology. In G. Morgan (Ed.), Organizational research strategies. Cambridge, MA: MIT Press.
- Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.
- Festinger, L. A. (1957). A theory of cognitive dissonance. Evanston IL: Row Peterson.
- Gergen, K. J. (1973). Social psychology as history. Journal of Personality and Social Psychology, 26, 309-320.
- Gergen, K. J. (1978). Experimentation in social psychology: A reappraisal. European Journal of Social Psychology, 8, 507-527.

Hormuth, S. E., Fitzgerald, N. M., & Cook, T. D. (1985). Quasi-experimental methods for community based research. In E. C. Susskind & D. C. Klein (Eds.), Community psychology: Methods, paradigms, and applications. New York: Praeger.

Israel, J., & Tajfel, H. (1972). (Eds.), The context of social psychology: A critical assessment. London: Academic Press.

McGuire, W. J. (1967). Some impending reorientations in social psychology. Journal of Experimental Social Psychology, 3, 124-139.

McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. Journal of Personality and Social Psychology, 26, 446-456.

Moede, W. (1920). Experimentelle Massenpsychologie. Leipzig: Hirzel.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 17, 776-783.

Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. Journal of Experimental Social Psychology, 3, 113-123.

Rosenthal, R. (1966). Experimenter effects in behavioral research. New York: Appleton-Century-Crofts.

Tajfel, H. (1972). Experiments in a vacuum. In J. Isreal & H. Tajfel (Eds.), The context of social psychology: A critical assessment. (pp. 69-119). London: Academic Press.

Triplett, N. (1897). The dynamogenic factors in pacemaking and competition. American Journal of Psychology, 9, 507-533.

Vorwerg, M. (1962). Wilhelm Wundt über das Experiment in der Psychologie. Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig, Mathematisch-Naturwissenschaftliche Reihe, 11.

The Random Sampling of Experiences in situ.

Stefan E. Hormuth

Universität Heidelberg

Running Head: Experience Sampling

Work using the experience sampling method and on this article was facilitated by grants (Ho 905/1) from the German Science Foundation (DFG) to the author. The author thanks M. Csikszentmihalyi for making him familiar with the method, and M. Lalli and A. Endres for their help in developing the Heidelberg version of the ESM. S. Franzoi, E. Kirchler, and R. Emmons made useful suggestions during the preparation of this article. Author's address: Stefan Hormuth, Psychol. Institut, Univ. Heidelberg, Hauptstr. 47-51, D-6900 Heidelberg 1, Germany (W).

### Abstract

Over the past decade, a number of investigators has developed a methodology for the sampling of everyday experiences, the experience sampling method (ESM). Some of these developments took place independently, some built on each other. As diverse as some of the beginnings of the method are the theoretical needs from which they arose. ESM is especially useful for the investigation of interactions between person and situation variables. This paper reviews some of the methodological and theoretical sources for a method to sample everyday behavior. It describes the practical procedures involved, together with some procedural and technological alternatives. An assessment of the method discusses the quality of the data that are being collected, the possible uses of the method in areas other than those where it is currently being used, as well as its limitations that require further refinements and the continued empirical assessment of experience sampling.

The random sampling of experiences in situ.

Over the past decades, a number of psychologists has felt the need for a method that gets close to the ongoing everyday behavior of humans. For this purpose, they have sampled thoughts and experiences at random points in time over several days in a person's life. Typically, subjects in these investigations carry timers and report or rate their experiences by filling out questionnaires at the time of a signal.

Behavior is conceptualized by these researchers as an ongoing "stream of behavior", as Barker (1963) has called it. This need arose from different sources, for instance, the call for the ecological validity of behaviors studied, the attempt to understand behavior as being embedded in an ongoing stream, the increased interest in the interaction of situation and person variables, the necessity to generalize from laboratory settings to the real world, the wish for the extensive study of single individuals, or the attempt to study processes that are difficult to create in a laboratory setting.

Ecological validity is a concept introduced originally by Brunswik (1949). It refers to the occurrence and distribution of stimulus variables in the natural or customary habitat of an individual. A psychological method is ecologically valid to the extent that its stimulus variables are a representative sample of those in the individual's habitat. This principle, when combined with the call for the study of functional organism-environment relationships (Brunswik, 1952), requires that "... situational circumstances should be made to represent, by sampling or related devices, the general or specific conditions under which the organism studied has to function" (p. 30). Brunswik went on to point out that the extension of the sampling requirement from persons to situations has methodological implications especially as far as the number of variables included in a research design is concerned. Brunswik did not necessarily suggest that his

thinking would lead to the study of behavior in situ, that is, in the actual situation. Rather, the way he conceived of representative designs assumes a prior knowledge of the population of stimuli (Stapf, 1976). The random sampling in rather than from situations, as it is practiced with the experience sampling method, is one answer to the request for the ecological validity of the study of personality.

Studying behavior in the situation to obtain data on the distribution of psychological phenomena was the methodological concern of Barker (1969). As an operationalisation of Lewin's (1943) concept of psychological ecology, he studied the 'behavior setting' as an ecological, situational unit. The method developed for that purpose was the painstaking, close observation over long periods of time (e.g., Barker & Wright, 1951) which in its detail resembled almost a duplication of reality. His observations were going on over longer time periods and thus led to an increased understanding of specific behaviors as part of an ongoing 'stream of behavior' (Barker, 1963). This metaphor, in turn, is already known from James (1890) who wrote about the 'stream of thought'. The conceptualization of thinking or behavior as an ongoing stream requires a method that is able to follow a person over a longer time period rather than creating or observing only short isolated episodes. However, a method such as Barker's that tries to assess the totality of behavior is for most purposes inappropriate. Any method trying to capture behavior as an ongoing activity must reduce the information already at the point of sampling.

The major source of several theoretical and methodological developments in personality psychology within the last fifteen years, however, was the apparent failure of the field to predict behavior adequately from personality variables. This was pointed out by a number of critiques as, for instance, Mischel (1968). Whether this failure was attributed to the overpowering effect of situational variables over

personality variables, and thus treated as a theoretical problem, or as an inadequacy of methodology in personality psychology, it had major consequences for future strategies of research. Kenrick and Dantchik (1983) point out the rise of theoretical models emphasizing the interaction between person and environment. On the methodological side, it led to a renewed interest in the idiographic approach to the study of personality, i.e. the attempt to use information from individual cases. Both of these developments contributed to the need for methods that allow to capture the behavior of individuals in individual situations.

If personality traits are defined as the stability of behavior over situations, and are then found to be of low predictive value, one possible response is a greater concern with the heretofore unquestioned part of the definition, that is, the situational variable. If a trait is only predictive in some situations but not in others, different situations have to be studied to determine how a certain trait is expressed. One possible approach is to ask subjects whether a specific trait is applicable to only some or all situations. This approach can be found in the work of Bem and Allen (1974). The advantage here is that the whole universe of situations possible for one person can be considered. The disadvantage is that it is only being considered hypothetically and subjectively. Another possible approach tries to create classes of situations and describe verbally or create experimentally prototypes of these classes to which subjects are exposed. Examples for the former can be found in the work of Bem (1983), examples of the latter in the work of Endler (1983) and associates.

Such approaches to study the interaction between personality traits and situations can reduce the possible range of situations to a large degree. But the description or creation of situations by the researcher also reduces ecological validity. The question whether the situations

or classes or situations selected are relevant to the trait investigated is not answered through such a design. In addition, it may not only be the behavior in a situation, but already the selection of a given situation that is determined by a given trait, as has been suggested by Snyder (1981). Situations selected and created by the researcher cannot be selected or avoided by the person being studied. Emmons, Diener and Larsen (1984) investigated the choice of situations through the experience sampling method. Finally, situations arise out of the interaction of a multiplicity of variables and thus present constantly changing degrees of novelty. This aspect again can only be created by the researcher to a diminished degree.

Several researchers considered the study of behavior in situations as they naturally occur the appropriate approach in face of the theoretical postulate of person-situation interaction. For instance, Pawlik and Buse (1982; Buse and Pawlik, 1984) were critical of methods that symbolically presented situations verbally and assessed the reaction to it by using questionnaires in cases where the registration of behavior and the access to natural situations is required. Desirable are, so Buse and Pawlik (1984) "ecologically valid studies, based on a representative sample of situations from the everyday life of subjects" p. 47, translated by the author). To this extent, they had developed a methodology and technology where behavior could be registered by the subject in everyday situations. In a similar vain, Schuster, Murrell, and Cook (1980) were interested in an exploratory answer to the question "What are the relative contributions of individual traits versus situations in predicting behavior?" To answer this question, they considered a "knowledge base of established relative person/setting/interaction predictive strengths for different categories of behaviors" (p. 27) in certain environments necessary. Both Schuster et al. (1980) and Buse and Pawlik (1984) considered Barker's (1968)

behavior setting as the appropriate operational definition of a situation. To determine the situational and personality components of behavior and their interaction, the experience sampling method has also been employed by Diener, Larson and Emmons (1984), Kirchler (1984), and Savin-Williams and Demo (1983).

As another result of the discussion whether personality psychology is able to predict behavior from traits, another question has received renewed attention, namely the right balance between a methodological approach that studies many individuals on one trait or many traits on one individual, as Stern (1921) described it, or the idiographic versus nomothetic approach, as Allport (1937) called it. In an updated survey of the different techniques for the study of individuals, Runyan (1983) includes a class of 'quantitative descriptive' methods which include several techniques that report the frequencies of certain behaviors over certain periods of time. Among them, and thus closely related to the experience sampling method, is time-budget research (Sorokin & Berger, 1939; Robinson, 1977) and Barker's methods of ecological observation.

Clearly, experience sampling shares some of the features of methods in this class. However, the differences between the time-budget method and the experience sampling method have to be stated clearly. While the former studies the relative amount of total time spent on certain activities, the latter takes random samples at certain times and records the activities taking place at that moment. Experience sampling yields large amounts of data on individual subjects. However, it is only rarely used as a single case method. Rather, since the method collects large amounts of data on several individual subjects, it allows for extended analyses of individual cases as well as for the useful aggregation over cases, thus possibly bridging a gap between the two methodological approaches of personality psychology. It may, however, also been usefully employed in single-case designs.

Naturalistic methods are frequently used to test results that had been obtained under controlled laboratory conditions in the real world and to extent the realm of these findings. The experience sampling method has been found useful for this purpose, for instance, in the area of self-awareness research (Duval & Wicklund, 1972). Csikszentmihalyi and Figurski (1982) analyzed the direction of subjects' attention using the experience sampling method and could relate it to the mood as reported at the same time. Their findings provided not only an extension of existing findings into the real world but contributed to the theoretical discussion as well. Franzoi (1984) also used the method to determine the occurrence of self-awareness, its affective concomitants in real-life situations, and its relationship to the trait of self-consciousness. Klinger, Barta, and Maxeiner (1980, preliminary study) used experience sampling of thought content to validate a questionnaire.

Under other circumstances, researchers wish to study events that are difficult to create under experimental or quasi-experimental conditions but involve naturally occurring events with implications that can only be studied by using records of subjects' everyday lives. For instance, Hormuth's (1983, 1984b; cf., Hormuth, 1984a) research on relocation as an opportunity for self-concept change involves the study of people's social behavior in a new environment. Reis, Wheeler, Spiegel, Kernis, Nezlek, and Perri (1982) were interested how a person's physical attractiveness determines that person's participation in everyday social life.

A different point of departure, namely to capture the ongoing stream of behavior or thought in its natural sequence and occurrence motivated, for instance, Hurlburt (1979) and McAdams and Constantian (1983) to develop respectively use methods that allowed to conduct investigations concerning "sampling of ongoing thought in subjects over

an entire day" (Singer, 1975, p. 734). McAdams and Constantian (1983) specifically used the experience sampling analysis to study the actualization of specific motives, namely intimacy and affiliation, in everyday life. Similar were the research interests of Klinger, et al. (1980), who studied the motivational correlates of thought content in a questionnaire study and used thought sampling as a validity check. Klinger had initially employed random thought sampling in 1978.

#### Methodological developments

Methodological developments toward the experience sampling method originated from several researchers. Because these developments sometimes took place without being aware of (or citing) each others work, it is not really possible to give credit to one individual or one group. The ordering in which these approaches will now be described is not meant to reflect the precedence of one approach over another.

Early studies using repeated self-report measures in everyday situations were frequently interested in mood and mood changes over the day. For instance, Nowlis and Cohen (1968) used hourly ratings on a mood adjective-check list to study the mood changes of three college students under academic pressure. Such studies are predecessors of the experience sampling method because they attempted to obtain data in the subjects' normal life at points in time that are independent of certain events and therefore constituted approaches at random sampling.

Hurlburt (1979) summarized the need for a technique that gathers information "(a) over relatively long periods of time; (b) in the subject's own milieu; (c) at times not contingent on any environmental event; (d) using a technique which minimizes reactivity; and (e) using a technique which minimally disturbs the individual's environment." For this purpose, he had a shirt-pocket sized random interval generator designed that emitted a 400-hz signal over an earphone at random intervals with a mean of 30 minutes. This device was worn by subjects

during all waking hours of the day. At the time of the signal, they had to write down in a 3 x 5 spiral notebook (a) their current thought; (b) what they were doing, and (c) the time of the day.

Hurlburt's method thus already involved random sampling of moments during the subject's everyday life. His primary interest was not in the events and behaviors taking place, but in the person's thoughts. His timer is similar to that still used by some other researchers, whereas the free notation of thoughts and events is unusual in that most other researchers provide questionnaires that are somewhat more prestructured.

Brandstätter proposed an experience sampling method in 1976 (Brandstätter, 1977, 1981, 1983). To allow the measurement of subjective well-being in everyday situations, he suggested the use of a random timer and self-reports of actual mood, current situation and suspected reasons for the current mood state over a four-week period. In his early actual research, he supplied participants with time tables or booklets where the times to report were already indicated. In later studies, he and his associates used watches that had to be set by the subjects themselves according to predetermined time tables. His subjects provided information in a booklet in free form. The content of the booklets was coded later by the subjects themselves.

A somewhat different approach was taken by Wheeler and Nezlek (1977) that should, however, be mentioned here, because it also attempted to use self-reports of everyday behavior, in this case social interactions, at the time the behavior takes place. Instead of a random sampling completely independent of certain events, Wheeler and Nezlek (and also later Reis, Wheeler, Spiegel, Kernis, Nezlek, and Perri, 1982) asked subjects to report on each interaction of ten minutes or more that occurs during a specified interval.

The probably most influential development of a technique for experiential sampling originated from Csikszentmihalyi and his

associates (Larson & Csikszentmihalyi, 1983). Csikszentmihalyi, using a radio-transmitted pager signal, employed experiential sampling on a great variety of research projects, for instance, on self-focused attention (Csikszentmihalyi & Figurski, 1982), on the effects of loneliness in adolescence (Larson & Csikszentmihalyi, 1978), or the influence of television on daily activities (Csikszentmihalyi & Kubey, 1981). While other investigators frequently only considered the method useful for specific purposes, or failed in working out a methodologically satisfactory and workable solution, Csikszentmihalyi and associates demonstrated the usefulness of the method to a wide variety of questions in a methodologically sound way. Their approach has also served as a model to other researchers, e.g., McAdams and Constantian (1983). The method as implemented by the current author and described here followed Ciskszentmihalyi's lead and tried to develop further technical and procedural improvements on that basis.

#### Procedure

The following describes the procedure involved in the use of the experience sampling method. Several variants to employ the method exist. The description here follows the procedure as used in Heidelberg by the author (Hormuth, 1983); where applicable, useful variations will be discussed.

The procedure uses as signalling device a timer that emits an audible signal (called the beeper) to remind the subject to fill out a questionnaire in a booklet. The beeper is carried by the subject for one week, during which eight reports are filled out daily.

The beeper currently used by this author has been especially constructed for use in the experience time sampling method. It is an electronic device, small enough so that it can be carried by subjects in a purse or pocket at almost all times, that is preprogrammed from a briefcase-size computer (Epson HX20) to elicit a signal at various

times. The size of the computer enables the researcher to prepare a beeper anywhere necessary, for instance, at a subject's home. The beeper allows for a maximum number of 128 calls and its battery life lasts more than eight days. That way, it can be given to the subject and need not be handled or recharged by anyone.<sup>1</sup>

During the week, subjects have to fill out a short questionnaire eight times daily at random intervals averaging 90 minutes. While for most subjects a schedule between 9 a.m. and 9 p.m. is appropriate, some subjects have different normal waking hours, for instance, persons working nights shifts. The computer program developed for this beeper makes it possible to have individualized schedules that take into account the subject's personal time table.

While the schedule uses random times to call on the subject, the times are not generated within the beeper. Such a random beeper, as it is used by several researchers does not provide the researcher with any control. It is useful to know the beep times so as to compare them with the records of the times kept by the subject. This is one way how the subjects' reliability in filling out the records can be assessed.

Twelve waking hours are determined for each day (usually 9 a.m. to 9 p.m.) that include working and leisure time. Within eight ninety-minute intervals per day, the beeper is supposed to emit a signal at a randomly determined point in time. Within each given ninety-minute interval, random times are generated, either by random table, and then accordingly typed into the computer program, or, in a variation on the previously used program, they are directly generated by the computer.<sup>2</sup> The computer provides a printout of each subject's individual schedule that can later be compared with the times as recorded by the subject. The individual schedule is then transmitted from the programming unit (the computer) to the beeper. It can then either be transmitted to other beepers as well, saved on file, or be replaced by another individual

schedule. Of course, enough flexibility is given for different research purposes and designs. For instance, the current schedule, with 64 beep times, uses only half of the capacity for 128 preprogrammed times.

The beeper is charged by four 1.5 volt batteries, lasting for approximately 8-10 days. Its size is 11 x 10.5 x 3.5 cm. Work is underway to further reduce this size. For the subject, it has an on-off "snooze"-switch to turn off the sound in situations where the subject does not want to be bothered. Subjects are required to note each such occasion and record the off times in a form (the "snooze report") and give a reason (e.g., "went swimming"). There is also a button to stop the beep sound. Subjects have no access to any other parts of the beeper. A program reset button is only accessible with a special tool. Experiences with a previous beeper have shown that subjects (or their friends) may try to open the beeper, take apart the electronic parts and even search for hidden microphones. A program reset is only accessible with a special tool.

Most researchers use a different kind of hardware for the experience sampling method. Csikszentmihalyi and his associates (cf. Larson & Csikszentmihalyi, 1983) use a pager that is set off via radio signals. It thus shares with the above described device the researcher's control over the timing of the signal. Currently, a pager may also be somewhat smaller and easier to carry because it is produced commercially in large quantities. However, the signal can only be transmitted within a 30 mile radius and may be disrupted by high rise buildings, hills, and other obstructions. Also, the signal has to be emitted at the real time and involves therefore more personpower. Especially the creation of individual schedules would be more difficult, if at all possible.<sup>3</sup>

The use of random timers (e.g., Franzoi, 1983; Klinger et al., 1980, Hurlburt, 1979) or wrist watches (Brandstätter, 1983; Diener and

associates, cf. Diener & Larsen, 1984) is also possible but for several reasons should only be considered a low-cost alternative. Random timers, as they are commercially available, are small devices that can be set to the average interval between signals. Thus, a setting of ninety minutes would have a similar effect as the schedule described above. However, the research participant has to be given some control over the device by turning it on and off at night hours. Because the signal is truly random, the researcher has no control and record of it. In addition, the battery life of approximately three days does not allow to sample for a whole weekly cycle without interference. Also not very satisfactory are other means of timing by which the subject cannot be kept unaware of the time at which the next recording can be expected when, for instance, wrist watches that have to be set by the subjects themselves. Randomness and the subject's unawareness seems to be an important factor in the application of the method.

As opposed, for instance, to random timers, an individually programmable beeper also allows for more flexibility in selecting certain times in a subject's day over others, and thus in more design variability. For instance, a study may concentrate on the distribution of certain activities, thoughts, or moods during work related activities, but at the same time try to obtain a smaller random sample of leisure time variables, serving as a control condition. It is then possible to program beep of a shorter average time apart during individual work times, and beeps more widely spaced during individuals leisure hours.

A complety different approach has been chosen by Pawlik and Buse (1982). Timing and registration is combined in a programmable microcomputer, specifically designed for experience sampling methodology. The computer is carried by the subject on the belt, and responses to questions on events and behavior are immediately typed into

the microcomputer by the subject. The data can then be directly transferred to another computer for analysis.

The beepers used in this author's research are designed to avoid any handling by the subject. That is, no recharging, loading of batteries, setting of times, or anything else should be necessary. One of the reasons for this hardware requirement is to keep subjects from perceiving any control over the beeper's functioning so as not to threaten the perceived random character of the signals. Also, it was found that being able to control one function of the device may lead to the attempt to control other functions as well. But the device does also need no handling by the research staff during the experience sampling period. This allows that the subject can move around freely without having to have any contact to the research staff and may even travel anywhere. This freedom of movement, not possible when paging devices are used, potentially allows the study of select, highly mobile or widely spread populations. Beepers, booklets, and instructions can easily be distributed and returned by mail.

The subject receives together with the beeper a recording booklet with instructions to fill out a questionnaire in the booklet immediately after the beep sound is heard or as soon as possible thereafter. The booklet, 10 x 15 cm, fits easily in a pocket. It contains 70 questionnaires of two pages each. Coding keys for locations and situations are in the front flap, the form for the "snooze report" attached to the back flap. The content of the questionnaire as well as some of the coding keys vary according to the purpose and population of the individual study. In the following, some of the principles we follow in constructing questionnaires for the experience sampling method are discussed.

As in any questionnaire construction, a basic decision has to be made about the degree to which the items and modes of answering have to

be prestructured. The problem, however, is exaggerated by the large amount of data on the one hand and cost and effort necessary to collect these data on the other hand.

Brandstätter (1983) and Hurlburt (1979) let their subjects take free notes that were later coded, either by the subjects themselves (Brandstätter) or by the investigator (Hurlburt). The free form answering mode has a major advantage for experience sampling in situ because it can accommodate the unexpected. It seems nearly impossible in the sampling of everyday experiences for the researcher to anticipate the possible situations, experiences or thoughts that may occur to a person.

These advantages, however, seem to be offset when the amount of data collected in the course of a week or several days is taken into account. There are more than fifty measurement times per subject. This author's studies involve typically around one hundred subjects. In a currently ongoing study, subjects carrys the beeper three times for one week each, spread over eight months. The rating of over five or even fifteen thousand answers is very time and labor consuming. On the subject's side, writing down answers in such a manner requires more time than a check on several items. Because the time needed to answer at any measurement point should be preferrably short (not more than two minutes seems advisable), only very little information could be collected that way. A short time requirement seems necessary to avoid interruption of ongoing activities as much as possible. Also, it is more likely that the questionnaire will be filled out at the required time if subjects do not have to draw attention to themselves in public--a frequently mentioned reason for failing to answer on time.

The cost of the method is another consideration in the decision how many data points should be collected at a single measurement point. The cost of the method is based on sampling, instrumentation, and time

needed for data collection. Sampling for participation in an involving and relatively long-term study has to take into account a refusal rate relatively high in relation to the number of persons contacted and high cost in overcoming this refusal. Instrumentation includes the cost of the electronic hardware and possibly the staff needed to operate it. The time needed for data collection, in this author's designs typically one week, means that the time needed to run one study can--dependent on the availability of hardware--be up to several months. Time will also usually need to be offset for the subject by an adequate renumeration, even though this only seems to play a role in recruiting subjects but not after the data are collected.

As a result of the above considerations, this author prefers a pre-structured questionnaire that allows for a relatively large amount of data to be collected within a period of approximately two minutes by a subject with some previous training. A questionnaire includes around twenty five items. In addition, the date and time have to be recorded by the subject. When compared to the subject's schedule, this provides one type of information on the reliability of filling the questionnaire out on time.

Those items that describe the type of activity, the current location, and the kind of person one is with are coded according to a coding system. It is attached to the booklet and can be folded out. It provides fifteen to twenty keys each for types of locations (e.g., at home, in another private home, at the place of work, etc.), types of company (e.g., alone, partner, member of household, neighbour, etc.), types of activities (e.g., studied, worked, on the way to something, slept, etc.), and types of conversation (e.g., the weather, food and drink, politics, friend, etc.). This way, subjects can describe their current activity very precisely on four items. While the combination of these coding keys encompasses a very large number of possible situations

to which a person can be exposed (in a current study 13 types of locations x 12 types of company x 22 categories of thought x 19 categories of conversation topics = 65,208 different situations), it may be advisable to tailor the system to the specific populations being studied. Some locations are unlikely for non-students, some activities may be unlikely for students. On the other hand, the flexibility of the system allows its use in very different studies and provides for their increased comparability.

The other items vary from study to study. They are either answerable on five-point scales or dichotomously (yes/no). Items refer either to the current state, that is, a description of the moment the beeper went off, or to the last time a event of a certain class took place before the signal. Examples of the former can be ratings of the current mood state or the direction of the focus of attention. These are continuous states that are captured at a random point in time. On the other hand are discrete events that may not have necessarily taken place at the time of the signal but are nevertheless of interest for the investigation. For instance, social interactions of different varieties occur frequently but are broken by times of being alone and without contact with others. In this case, answers refer to the last social interaction that took place between the last and the current signal.

Of some sampled events, again only a sample can be described. For instance, if the last social interaction has been with a group of people, and the characteristics of one partner or conversation are of interest, the subject has to be supplied with simple decision rules to select only one partner from the group and answer the following questions in reference to this one interaction partner. The most simple decision rule is to select the last conversation partner before the signal. Similar decisions have frequently to be made when several events of a similar nature are co-occurring and one has to be selected as

the target for several items on the questionnaire.

The actual procedure involves recruiting, training, and experience sampling. Recruiting subjects is, of course, more determined by the research purpose and population of interest than by the method chosen. However, the researcher has to be aware that some specific conditions apply. In particular, the method appears at first to the potential subject to be demanding in terms of time and interruption of the daily schedule. This is only true to some degree but may lead many contacted persons to reject the idea of participation outright. A method of recruiting, carefully tailored to the target population, together with financial or other incentives, has here to be considered part of the research design. Because of the sometimes interruptive character of the method, the decision about participation or continuation may also depend on supervisors, school principals and other persons that would not be affected by other research methods. In this author's research, studying adaptation to new environments on general (non-student) populations, the rate of participation is typically around 20%. For an adolescent population, Larson and Csikszentmihalyi (1982) report a much higher participation rate of 54%. Comparable data could not be found in the reports of other authors.

Once subjects are recruited, a training phase will usually have to be arranged. The task to assess a situation on many different dimensions, using a variety of very differentiated coding systems, is unusual to lay persons and sometimes difficult. Without prior training, subjects may make use of only a small subset of the available coding keys, or consider the task too difficult and comply less. The amount of training needed varies with the complexity of the task. Few open-ended items require less training than a coding system. Very sophisticated technology, as Pawlik and Buse's (1982) fully computerized event registration device where both questions and answers have to be coded,

cannot be usefully implemented in research without training subjects very carefully. Pawlik and Buse designed a training, lasting for a six-day period, that consists of video presentations and real life situations, and stresses the rating reliability of subjects.

In this author's research, the actual training consists of a two-hour introduction to the method and the coding scheme and one trial day that should also facilitate the adaptation to the beeper. After the end of the trial day, possible problems with the booklet or the beeper can be discussed before the actual one-week period of experience sampling starts the next day. As opposed to Pawlik und Buse, this training stresses familiarity with the method and the coding keys and does not try to make trained raters out of subjects. The degree of familiarity with the coding scheme required and the relative emphasis on the objective assessment of situational characteristics versus subjective perceptions has to determine for the individual researcher the kind and amount of training necessary. Where the interaction of situations and personality variables is of interest, a reliable assessment of the situation may be necessary without precluding the concurrent rating of the subjective experience of this particular situation.

After a careful training, the actual experience sampling requires little staff involvement. It is useful to provide subjects with a phone number where they can always ask for help in case of problems. These problems can involve equipment failure (due, for instance, to faulty batteries), sudden illness that requires bedrest and thus limits greatly the range of experiences that can be sampled, and objections by superiors (or, as in one case, religious objections) to participation. In this author's studies, involving several hundred weeks of experience sampling, less than 3% of all subjects had to discontinue participation for these or other reasons.

At the end of the experience sampling period, an evaluation takes

place. In a short questionnaire (Table 4), subjects are asked for their experiences with the beeper device, with the questionnaire, the degree to which the method disrupted or influenced their regular activities, and their self-perceived accuracy in filling out the questionnaire on time and completely. A similar evaluation was done by Pawlik and Buse (1982). Little information can be gained that way about the actual quality of the data (although some authors report such information for this purpose), however, some suggestions for necessary improvements can be taken from there.

#### Some data analysis examples

There is no commonly agreed upon way for the analysis of data collected through experience sampling. The guiding principle for most investigators is the useful reduction of the large amount of data typically obtained. The different approaches are determined by different research objectives as, for instance, the stability of behavior over situations or over different states of the individual.

If different categories for states (e.g., self-aware or not) or situations (e.g., with or without another person) are reported, an analysis of the relative frequencies of all categories can eliminate from a large number of potential categories those with zero frequencies and allow for the combination of some with moderate frequencies. Most investigators proceed to calculate aggregate scores, summarizing either over persons, over situations, or over persons in situations. For instance, Brandstätter (1981) reports average specific mood ratios for four categories of locations where persons were, four types of activities they engaged in, and the types of social situations the person was in.

The most commonly employed statistical analyses of experience sampling data are simply correlations and analysis of variance. Correlational analyses can provide information about the co-occurrence

of, for instance, certain states, such as self-awareness, and other states, such as affect, or activities, such as work or social behavior. For instance, Franzoi and Brewer (1984) tried to determine the degree of self-awareness by asking subjects for the amount of time they had spent thinking about themselves. The several hundred self-reports obtained were correlated with evaluation and activity ratings at that point in time. Franzoi and Brewer adjusted for the effects individual subjects may have on these relations by partialing individual subject variations out and calculated partial correlation coefficients.

Correlational analyses are also used to analyze the relationship between state variables, as measured using the experience sampling method, and trait variables obtained independently of the experience sampling. Franzoi and Brewer also were interested in the relationship between trait variables, namely self-consciousness, and the state variables of self-awareness. Subjects' ratings of amount of time spent in self-awareness were summed and averaged across all time intervals, yielding mean self-awareness scores for each respondent. They obtained thereby one self-awareness (state) score for each subject that was correlated with the (trait) measure of self-consciousness.<sup>4</sup>

An investigation by Buse and Pawlik (1984) can serve as an example of the use of analysis of variance to test for intersituational inconsistency. Their analyses adapted Golding's (1975) considerations to the specific characteristics of experience sampling data. They had tested the consistency hypothesis by way of inter-setting correlations, that is, the correlations of state measurements, for instance mood items, between seven classes of settings. They also provided an item discrimination index for each situation and as a global measure for inter-setting consistency Cronbach's alpha. These analyses allow to determine the variance explained by person factors across situations. They do not, however, determine the situational variance, that is, the

inconsistency of behavior across situations, because of the measurement error contained in the rest variance. In an analysis of variance, the interaction between persons and settings provides the situation specific variance that cannot be explained by general tendencies such as traits. The factors of their analysis of variance design were (A) settings (using only four levels), and as personality traits (B) neuroticism, and (C) extraversion. Thus, the ANOVA was a  $4 \times 2 \times 2$  factorial design with repeated measurement on factor A. Dependent variables were the relative frequencies of 82 behavioral items, thus yielding 82 different ANOVAs. Buse and Pawlik differentiate between two types of interactions. Type I is only related only to the relative size of the difference traits make between behavior in settings. Type II also influences the direction of differences. A Type I interaction does not influence the rank order of personality types whereas a Type II interaction does.

A different set of analyses was used by Schuster et al. (1980). Their analyses were conducted at two levels: one on the individual level, one on the aggregate level. The first set of analyses studies "individual subjects in each setting to determine whether a given child's behavior in a particular setting more closely resembled the behavior of the other children in the same setting or that of his/her own in other settings. The aggregate analysis provided an overall comparison of the variation produced by person differences, by setting differences, and by the interaction of setting and person" (p. 30). Schuster et al. describe in reasonable detail maximum-likelihood procedures employed, namely a minimum chi-square procedure for goodness of fit on the individual level and a multidimensional contingency table analysis (Goodman, 1971a, 1971b) on the aggregate level.

The range of possibilities for the analysis of experiential sampling data seems to be far greater than that reported in the literature sofar. For instance, almost all analyses that can be found

in the literature concentrate on the co-occurrence of behaviors, states and situations, or the relation to external trait variables, while their succession and temporal sequence is rarely addressed. However, there are also some limits to such analyses and possible generalizations from experience sampling data, due to the autocorrelations between the events sampled (Brandstätter, 1981).

#### Assessment and discussion

An assessment of experience sampling and the quality of the data obtained through this method has to focus on the role of the research participant. The quality of the data collected by experience sampling depends almost exclusively on that person. Their reliability and validity can to some extent be determined by answers to the following questions: Does the subject respond to the signals on time? Are the subject's objective circumstances influenced by participation? Is the subject's subjective perception of a situation influenced by the method? Is the subject able to report on or rate a situation? Most of these questions lack still sufficient answers.

To preserve the random sampling of situations, signals have to be responded to on time. If this were not the case, subjects would wait for a convenient situation, that either allows more time or is more comfortable to report on than the randomly sampled one. The basic purpose of the method would thus be threatened. Therefore, control of the timeliness of responses seems essential.

The pre-programmed beeper provides printouts of the individual signal schedules provided for subjects. These can be compared with the time written down on each sheet by the subject at the time of responding. This information provides at least an indication whether the beeper was with the subject at that time. Its accuracy is one bit of information to help in assessing the reliability of the whole report at a given time. The following tables provide data on the timeliness of

responding to the signal, based on the time difference between the scheduled call and the reported time. Table 1 presents results on the signal level. Based on 5,145 signals, to which all subjects responded, 70% were answered within three minutes, 80% within five minutes. These seem reasonable response times. However, the 95th percentile is only reached after 43 minutes.

-----  
Insert Table 1 about here  
-----

Responding to each signal can depend on a great number of factors and may show intraindividual as well as interindividual variance. To determine whether subjects in general respond on time, the mean response time for each of 101 subjects was also calculated (Table 2). Here, ten percent of all subjects respond on the average within 1.3 minutes, 20% within two minutes, and 50% still within 4.5 minutes. These relatively low average response times on the subject level seem to indicate a willingness on the part of the research participants to comply with the signal. Sometimes situational demands interfere, hence the greater time differences on the signal level.

-----  
Insert Table 2 about here  
-----

An analysis of experience sampling reports should, if possible, compare those reports answered within a short time period with those answered only after a while. If these reports on situations differ, it may either be due to situational characteristics that made an immediate response difficult, or be due to the fact that the situation responded to was not selected randomly anymore but rather by the subject. The exclusion of such reports from the general data analysis has to be decided by the investigator based on other considerations as well.

Some signals are never responded to. The data above were collected in a study that provided a schedule with eight signals on each of eight days. All 64 signals were only responded to by one subject. Fifty percent of the subjects provided more than 53 reports, that is, a median response quote of 82%, and 90% still more than 38 (Table 3). These response frequencies seem to be comparable with those obtained by other investigators, as far as reported. Pawlik and Buse (1982) report an average response to 57 out of 66 signals, that is an average response quote of 86%. As other examples, Csikszentmihalyi and Figurski (1982) obtained an average response rate of 80% (45 out of 56 on the average), and Savon-Williams and Demo (1983) 78%.<sup>5</sup>

-----  
Insert Table 3 about here  
-----

Also of some interest is the research participants' perception how they were able to comply with the demands of the experience sampling method. A post-sampling questionnaire was administered. Overall, it seems that subjects thought the experience sampling to be not very difficult and not very disruptive. However, the immediate response to the signal is frequently not possible. That subjects did not think all their daily activities were adequately included in the questionnaire is appropriate since this was not the purpose of the investigation. Encouraging is the high willingness to participate in such a study again, as expressed not only on this questionnaire but in personal conversations as well. Similar positive responses by subjects in experience sampling are reported by Pawlik and Buse (1982).

-----  
Insert Table 4 about here  
-----

Generally, however, data on compliance are rarely found. In no

case was an analysis undertaken where the quality of different data was compared based on such criteria. Considering the importance of the random element in the method, more attention should be paid to the timeliness of reports. In some situations, it may indeed be better to instruct the subject to omit a report rather than to complete it fourty-five minutes after the signal.

The stability of self-ratings over the experience sampling period can be influenced by the subjects' carefulness, but also by adaptation to the beeper and the self-report of experience. One way of approaching the question whether subjects' adaptation to the method influences their rating reliability is to compare the ratings during the beginning of the rating period with those toward the end. This tests for a disruption that can be adapted to. Pawlik and Buse (1981) compared the first three days of a six-day rating period with the second half. The median correlation coefficients ranged from  $r = .57$  (locations) to  $r = .76$  (mood) and  $r = .80$  (motives). The individual correlations were lowest for those rating categories rarely needed. Overall, so Pawlik and Buse, the stability of individual items compares favorably to that of questionnaire items. Pawlik and Buse had subjects become familiar with the method in a training period. This author, too, considers a training period of at least one full day necessary. Based on these stability data, there is no reason to assume that increasing familiarity with the beeper and the rating alters the ratings over time.

It is more difficult to assess those influences on everyday life that are constant or those situations that cannot be included in the experience sampling. Both of these influences can affect the ecological validity of the method. Since most other methods are more vulnerable to threats to ecological validity than the one discussed here, including observational methods, no one comparison with other types of data can be used to assess this type of validity. Pawlik and Buse instructed

subjects prepare every evening a list of settings where they had spent the day, in addition to having them recorded at the time of the signal. They found that most unresponded to signals occurred in the home (where their relatively heavy signal and registration device may have been left sitting on a table) or at places, like discotheques, pools, and the gymnasium, where they had instructed their subjects not to take the somewhat vulnerable recording device. In this authors research, subjects are not given such instructions since the beeper is deliberately built to withstand heavy stress. Rather, subjects are given the opportunity to turn the beeper off if they so desire and to fill out a form (the "snooze" report), indicating the times and reasons. These self-reported off-times include situations where the device could not be carried, as at the pool, and a few public situations where the sound could be considered disruptive, as in a church.

The subjective perception of the situation could also be influenced by a method that is based on self-observations. Several authors speculate on the relationship between the method and self-awareness (Duval & Wicklund, 1972) as it may be induced by the necessity for self-observation. Brandstätter (1984), for instance, assumes that this may be improving the accuracy of the self-reports. Indeed, the self-awareness literature gives some indication that the validity of self-reports is increased by self-awareness that is created in addition to the already existing demands of a self-report questionnaire (Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977; Gibbons, 1983). On the other hand, self-awareness can have behavioral effects, such as facilitation of responses, conformity (cf., Wicklund, 1975), or the adherence to internal standards (Hormuth, 1982) that would change the experience of the sampled situation.

However, the degree to which experience sampling may raise self-awareness is difficult to determine. Wicklund (1975) already raised

general doubts whether self-awareness can be assessed in a manipulation check since the question in itself, in turn, creates some self-awareness. Csikszentmihalyi and Figurski (1982), Franzoi and Brewer (1984), and Hormuth (1984b) specifically tried to assess the degree of self-focused attention subjects engaged in. As with other data, a comparison base level using ecologically valid data is not available. Csikszentmihalyi and Figurski obtained self-awareness information rating the open-ended answers to the question: "What were you thinking about when you were beeped?". Using a rather narrow definition of self-related thoughts, about 8% of thought content was rated as indicating self-awareness. Hormuth (1984b) asked subjects to rate on a dichotomous item whether their attention was rather focussed on the self or on the outside. In 31% of all ratings, subjects reported self-awareness. Franzoi and Brewer (1984) asked subjects how their focus of attention during the last ten minutes before the signal was divided (in minutes) between private aspects of the self, public aspects of the self, and other people or things. Pooling the results of both of their studies with a total of 84 subjects, subjects reported to have engaged in private self-awareness for about 4.1 out of the ten minutes, in public self-awareness for about 2.6 minutes, and in thoughts on other persons or things for about 4.4 minutes. While Csikzentmihalyi and Figurski may underestimate the actual ratio of self-awareness, Franzoi and Brewer may have pointed the subjects' attention to the self by their instructions and thus overestimate the amount of time engaged in self-awareness. All three sets of results, however, can only describe the ratio of self-related to non-self-related thoughts during the experience sampling period without any basis for comparison.

If experience sampling raises the general level of self-awareness, it may also influence chronic self-awareness, because sampling periods typically last for several days. Chronic self-awareness, usually

referred to as self-consciousness, can be assessed by the self-consciousness scale (Fenigstein, et al. 1975). Franzoi and Brewer gave this scale to one half of their subjects before and the other half after the experience sampling period. A general increase of self-awareness could be indicated by higher SCS scores after than before the sampling period. This was not the case. Because their sampling period was relatively short (two days), no data are available for longer times. Nevertheless, the available information suggests that overall self-consciousness is not raised significantly and self-awareness at the time of the signal is not the dominant focus of attention. Therefore, unless specific self-awareness manipulations or instructions at the time of the signal are included, the influence of the method does not seem to be enough to change the situation nor increase the accurateness of perception.

Experience sampling methodology has established itself as a tool to collect psychological data that may be inaccessible through other methods. It is especially useful to address some questions of great relevance to personality research, for instance, the interaction of person and situation variables, the conditions for the actualization of traits and motives in everyday behavior, the validation of questionnaires on ecologically valid data, and the choice of situations. While ESM has already been employed for a great variety of research issues in personality and other areas of psychology, far more seem possible, for instance, in the investigation of complex social behaviors, or the succession and dependence of situations over time. In idiography and clinical psychology, the establishment of base rates and behavioral changes through experience sampling can greatly enhance the usefulness of single case designs. In behavior therapy, random sampling may even prompt other behaviors besides the response to a questionnaire.

The current methodological weaknesses lie mainly in the

responsibility the research subject is given in collecting not only subjective data, like thoughts and feelings, but also objective ones, like the description of situations. Being left alone for long periods of time, and being untrained raters, the reliability of responses is difficult to determine. Investigators may want to obtain different kinds of data in addition to experience sampling, for instance, observational, peer ratings, retrospective, or through questionnaire. Also, controls and check&on the subject should be built into experience sampling designs. Checking on the timeliness of responses is only one such check, others will have to be devised. The collection of objective data and the use of coding systems requires careful training of research participants.

Most difficult is the assessment of the validity of the information obtained through the experience sampling method. Almost by definition, a comparison with data collected through other psychological methods cannot provide this information. To assure the ecological validity of the data, hardware improvements should continue to have the unobtrusiveness in everyday life in mind, so that signals can be received and reports be given in almost every setting. A theory-guided comparison with other types of data may also provide information on validity in individual studies.

Experience sampling is a rather costly method for all concerned, investigators and subjects. It has to be employed and further developed in such a way that the quality of data can be controlled for and, where possible, improved. On the other hand, the comparisons with some other methods, as in the extension of laboratory findings to the real world or the validation of questionnaires, has shown that the established methods of psychology can also provide useful, reliable and valid information. Some questions, however, cannot be answered by these established methods and there ESM enhances the choice of tools available to the researcher.

## References

- Allport, G. W. (1937). Personality: A psychological interpretation. New York: Holt.
- Atkinson, J.W. (1981) Studying personality in the context of an advanced motivational psychology. American Psychologist, 36, 117-128.
- Barker, R. G. (1963). The stream of behavior. New York: Appleton-Century-Crofts.
- Barker, R. G. (1969). Ecological psychology. Stanford: Stanford University Press.
- Barker, R. G. & Wright, H. F. (1951). One boy's day. New York: Harper and Row.
- Bem, D. J. (1983). Toward a response style theory of persons in situations. In M. M. Page (Ed.), 1982 Nebraska Symposium on Motivation: Personality--current theory and research. Lincoln: University of Nebraska Press.
- Bem, D. J. & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review, 81, 506 - 520.
- Brandstätter, H. (1977). Wohlbefinden und Unbehagen. In W. H. Tack (Ed.), Bericht über den 30. Kongreß der Deutschen Gesellschaft für Psychologie in Regensburg 1976. Göttingen: Hogrefe.
- Brandstätter, H. (1981). Time sampling of subjective well-being. In H. Hartmann, W. Molt, & P. Stringer (Eds.), Advances in economic psychology. Heidelberg: Meyn.
- Brandstätter, H. (1983). Emotional responses to other persons in everyday life situations. Journal of Personality and Social Psychology, 45, 871-883.

- Brunswik, E. (1949). Systematic and representative design of psychological experiments. Berkeley and Los Angeles: University of California Press.
- Brunswik, E. (1952). The conceptual framework of psychology. International Encyclopedia of Unified Science, Vol I, no. 10. Chicago: University of Chicago Press.
- Buse, L., & Pawlik, K. (1984). Inter-Setting-Korrelationen und Setting-Persönlichkeits-Wechselwirkungen: Ergebnisse einer Felduntersuchung zur Konsistenz von Verhalten und Erleben. Zeitschrift für Sozialpsychologie, 15, 44-59.
- Buss, D. M., & Craik, K. H. (1983). The dispositional analysis of everyday conduct. Journal of Personality, 51, 393-412.
- Cattell, R. B. (1957) Personality and motivation: Structure and measurement New York: World Book.
- Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. Journal of Personality, 50, 15 - 28.
- Csikszentmihalyi, M. & Kubey, R. (1981). Television and the rest of life: A systematic comparison of subjective experience. Public Opinion Quarterly, 45, 317 - 328.
- Diener, E., & Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. Journal of Personality and Social Psychology, 47, 871 - 883.
- Diener, E., Larson, R.J., & Emmons, R. A. (1984). Person x situation interactions: Choice of situations and congruence response models. Journal of Personality and Social Psychology, 47, 580 - 592.
- Duval, S., & Wicklund, R.A. (1972). A theory of objective self-awareness. New York: Academic Press.
- Emmons, R. A., Diener, E., & Larsen, R. J. (1984) An extension of two

models of interactionism: Choice and avoidance situations and congruence response models. Manuscript under review, University of Illinois at Urbana-Champaign.

Emmons, R. A., Diener, E., & Larsen, R. J. (1984). An affect-goal analysis of situational choice behavior. paper presented at the 92nd annual convention of the American Psychological Association, Toronto, Canada.

Endler, N. S. (1983). Interactionism: A personality model, but not yet a theory. In M. M. Page (Ed.), Nebraska Symposium on Motivation 1982: Personality - current theory and research. Lincoln: University of Nebraska Press.

Endler, N.S., & Magnusson, D. (Eds.). (1976), Interactional psychology and personality. New York: Wiley.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. Journal of Consulting and Clinical Psychology, 43, 522 - 527.

Franzoi, S.L., & Brewer, L. C. (1984). The experience of self-awareness and its relation to level of self-consciousness: An experiential sampling study. Journal of Research in Personality, 18, xxx - xxx.

Gibbons, F. X. (1983). Self-attention and self-report: The "veridicality" hypothesis. Journal of Personality, 51, 517 - 542.

Golding, S. (1975). Flies in the ointment: Methodological problems in the analysis of the percentage of variance due to persons and situations. Psychological Bulletin, 82, 278 - 288.

Goodman, L. A. (1971a). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models of multiple classifications. Technometrics, 13, 23 - 61.

Goodman, L.A. (1971b). Partitioning of chi-square, analysis of marginal contingency, and estimation of expected frequencies in multidimensional contingency tables. Journal of the American Statistical Association, 66, 339 - 344.

Hormuth, S. E. (1982). Self-focused attention and the adherence to personal standards. In K. Obuchowski & A. Kossakowski (Eds.), Progress in psychology of personality (pp. 95 - 105). Berlin, GDR: VEB Deutscher Verlag der Wissenschaften.

Hormuth, S. E. (1983). Ortswechsel als Gelegenheit zur Änderung des Selbst. (Relocation as an opportunity for self-concept change). Report and continuation proposal to the German Science Foundation (DFG). University of Heidelberg, Psychological Institute.

Hormuth, S. E. (1984a). Transitions in commitments to roles and self-concept change: Relocation as a paradigm. In V. L. Allen & E. van de Vliert (Eds.), Role transitions: Explorations and explanations. New York: Plenum.

Hormuth, S.E. (1984b). An ecological model of self-concept change: Relocation as a paradigm. Unpublished data. University of Heidelberg, Psychological Institute.

Hurlburt, R. T. (1979). Random sampling of cognitions and behavior. Journal of Research in Personality, 13, 103 - 111.

Kenrick, D.T., & Dantchik, A. (1983). Interactionism, idiographics, and the social psychological invasion of personality. Journal of Personality, 51, 286-307.

Kirchler, E. (1984). Das Befinden von Wehrpflichtigen in Abhängigkeit von personellen und situativen Gegebenheiten. Zeitschrift für Arbeits- und Organisationspsychologie, xx, yyy - zzz.

Klinger, E. (1978). Dimensions of thought and imagery in normal waking states. Journal of Altered States of Consciousness, 4, 97 - 113.

- Klinger, E., Barta, S. G., & Maxeiner, M. E. (1980). Motivational correlates of thought content frequency and commitment. Journal of Personality and Social Psychology, 39, 1222 - 1237.
- Larson, R., & Csikszentmihalyi, M. (1978). Experiential correlates of time alone in adolescence. Journal of Personality, 46, 677-693.
- Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. Reis (Ed.), New directions for naturalistic methods in the behavioral sciences. San Francisco: Jossey-Bass.
- Lewin, K. (1943). Defining the 'field at a given time'. Psychological Review, 50, 292 - 310.
- McAdams, D., & Constantian, C.A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. Journal of Personality and Social Psychology, 45, 851-861.
- Milardo, R. M., Johnson, M.P., & Huston, T.L. (1983). Developing close relationships: Changing patterns of interaction between pair members and social networks. Journal of Personality and Social Psychology, 44, 964-976.
- Mischel, W. (1968). Personality and assessment. New York: Wiley.
- Nowlis, V.P. & Cohen, A.Y. (1968). Mood-reports and the college natural setting: A day in the lives of three roommates under academic pressure. Psychological Reports, 23, 551-556.
- Pawlik K., & Buse, L. (1982). Rechnergestützte Verhaltensregistrierung im Feld: Beschreibung und erste psychometrische Überprüfung einer neuen Erhebungsmethode. Zeitschrift für Differentielle und Diagnostische Psychologie, 3, 101-118.
- Pervin, L. A. (1976). A free-response description approach to the analysis of person-situation interaction. Journal of Personality and Social Psychology, 34, 465-474.

- Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R., & Hood, R. (1977). Self-focused attention and self-report validity. Journal of Personality, 45, 513 - 527.
- Reis, H. T., Wheeler, L., Spiegel, N., Kernis, M. H., Nezlek, J., & Perri, M. (1982). Physical attractiveness in social interaction: II. Why does appearance affect social experience? Journal of Personality and Social Psychology, 43, 979 - 996.
- Reis, H.T., Wheeler, L., Spiegel, N., Kernis, M.H., Nezlek, J., & Perri, M. (1982). Physical attractiveness in social interaction: II. Why does appearance affect social experience? Journal of Personality and Social Psychology, 43, 979-996.
- Robinson, J. (1977). How Americans use time. New York: Praeger.
- Runyan, W. M. (1983). Idiographic goals and methods in the study of lives. Journal of Personality, 51, 413 - 437.
- Savon-Williams, R. C., & Demo, D. H. (1983). Situational and transituational determinants of adolescents self-feelings. Journal of Personality and Social Psychology, 44, 824-833.
- Schuster, S. O., Murrell, S.A., & Cook, W.A. (1980). Person, setting, and interaction contributions to nursery school social behavior patterns. Journal of Personality, 48, 24-37.
- Snyder, M. (1981). On the influence of individuals on situations. In N. Cantor and J. F. Kihlstrom (Eds.), Personality, cognition, and social interaction. Hillsdale, NJ: Erlbaum.
- Sorokin, P. & Berger, C. (1939). Time budgets of human behavior. Cambridge, MA: Harvard University Press.
- Stapf, K. H. (1976). Bemerkungen zur Gegenstands- und Methodendiskussion in der Umweltpsychologie. In G. Kaminski (Ed.), Umweltpsychologie: Perspektiven, Probleme, Praxis. Stuttgart: Klett.

- Stern, W. (1921). Die differentielle Psychologie in ihren methodischen Grundlagen. (Differential psychology in its methodological foundations.) Leipzig: Barth.
- Wheeler, L., & Nezlek, J. (1977). Sex differences in social participation. Journal of Personality and Social Psychology, 35, 743-754.
- Wicklund, R. A. (1975). Objective self-awareness. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, Vol. 8. New York: Academic Press.
- Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. Journal of Personality and Social Psychology, 43, 111 - 122.

Footnotes

- 1) The beeper and the accompanying software were designed and built resp. written by Dipl.-Ing. Wolfgang Kuck, Rosenheim, Bavaria, FRG, according to this authors' instructions.
- 2) This variation of the software was programmed by Dipl.-Psych. Marco Lalli.
- 3) It was for some of these reasons that the original suggestion for a self-contained device was made by M. Csikszentmihalyi to me, for which I am thankful (Csikszentmihalyi, personal communication, Summer 1981).
- 4) Franzoi and Brewer (1984) studied both the private and public aspects of self-awareness and self-consciousness. This aspect is not needed for the present argument and is omitted here for ease of presentation.
- 5) Some of the missed signals are due to equipment failure when prototypes of the beepers were used. The beepers now used by this author are more reliable. The response rate reported by Csikszentmihalyi also includes some equipment failure.

Table 1: Timeliness of responding to the signal based on individual signals.

Minutes between signal and reported time	Cumulative Percent of signals reported within time
1	50
2	60
3	70
4	75
5	80
8	85
18	90
43	95

Note: Unpublished data from Hormuth (1984), Study IV. N = 5,145.

Table 2: Timeliness of responding to the signal based on the average response time of individual subjects.

Mean number of minutes between signal and reported time	Cumulative percent of subjects reporting within time
less than 1	5
1.3	10
2.1	20
2.9	30
3.7	40
4.6	50
5.4	60
7.1	70
8.7	80
11.7	85
15.4	90
28.8	95

Note: Unpublished data from Hormuth (1984), Study IV. N = 101.

Table 3: Percentage of signals responded to over the whole sampling period of eight days.

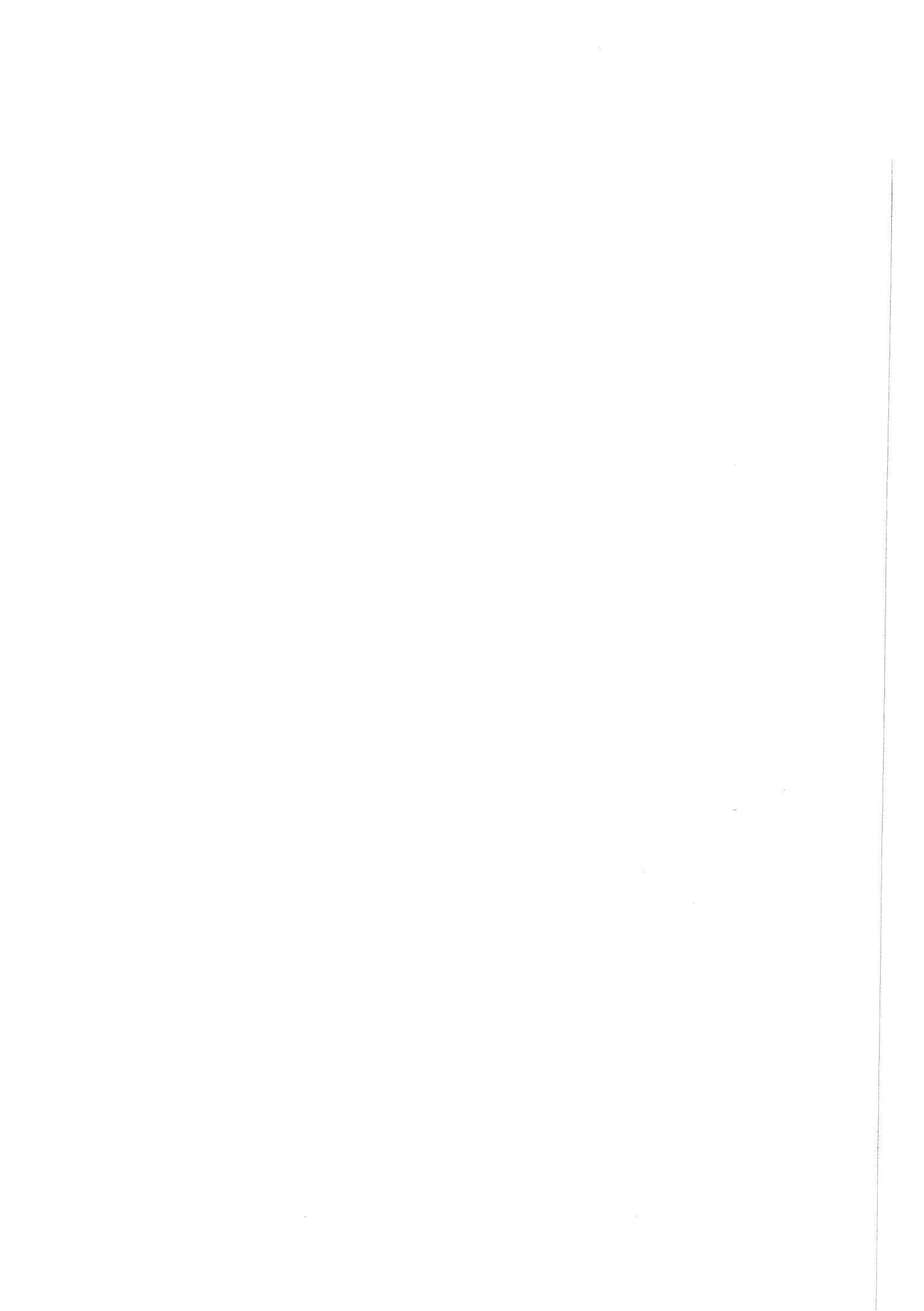
Percentage of responses (out of 64 possible)	Percentage of subjects (N = 101)
47	95
59	90
67	80
83	50
90	25
95	15

Note: Unpublished data from Hormuth (1984), Study IV.

Table 4: Evaluation by research participants after the experience sampling.

Item	Percent yes-answers
I sometimes put the beeper somewhere and forgot it there.	24
The beep is bothering in public.	14
I sometimes forgot to turn the snooze-function back on.	3
The questionnaire is too long.	5
Filling it out takes too much time.	9
Filling it out disrupts the daily routine.	22
It is frequently not possible to fill out the questionnaire immediately.	65
Too much time passes between the signal and filling out the questionnaire.	12
The response categories are inappropriate.	34
The questionnaire assesses my day well.	23
I believe that some important activities are not all taken into account.	20
I would participate again in such a study.	75

Note: Unpublished data from Hormuth (1983), Study IV. N = 101.



The Role of Urban Environments for the Self-Concept:  
A Research Example Using an Autophotographical Approach

Stefan E. Hormuth & Marco Lalli

Universität Heidelberg

Paper presented at the 8th International Conference on "Environment and Human Action" of the International Association for the Study of People and their Physical Surroundings (iaps 8), Berlin (West), July 25 - 29, 1984. This research was supported by the German Science Foundation: DFG Ho 905/1-1. Address correspondence to: Dr. Stefan Hormuth, Psychol. Institut, Universität Heidelberg, Hauptstr. 47-51, D-6900 Heidelberg 1, Germany/W.

## The role of urban environments for the self-concept

A person's understanding of self is acquired and developed through social experience. This basic idea has guided almost all theory and empirical research on the self-concept in modern psychology, beginning with William James in 1890, elaborated upon by Cooley (1902) and George Herbert Mead (published in 1934), and tested empirically with renewed emphasis within the last decade or so, as reviewed by Shrauger and Schoeneman (1979).

The conception of the role of urban environments for one's self which serves as the basis for the research to be presented here makes use of this idea. Therefore, let me first try to draw together several lines of thinking about the sources and symbols of social experience. The constituents of the self and the objects that mediate and perpetuate social experience will be described as the ecology of the self, and urban environments are seen as one integral part of this ecology of the self. The self both shapes this ecology and is a reflection of it. The constituents of the ecology of the self are others, as the sources of direct social experience, things, as symbols and representations of social experiences, and environments, as the setting for social experiences. They are reflected and organized in self-related cognitions.

The first function of others for the concept of self is the reflection and reaction to one's actions, both verbal and nonverbal. These ideas are rooted in the writings of Cooley (1902) and Mead (1934). The reactions of others to one's own actions provide a mirror for the self. Both the self and societal rules, expectations, and reactions thus depend upon each other and are unthinkable without the other. Exchange with others is therefore a process by which the self can be maintained as well as changed.

Others do not necessarily have to be physically present to provide social experiences nor do they have to be existing as a concrete person.

Social experiences are symbolized in various forms, for instance, in rules and expectations, but also in symbols and objects. Things have several functions for the self, as, for instance, described by Csikszentmihalyi & Rochberg-Halton (1981) or Graumann (1974). For many social experiences they provide the necessary tools, such as a record player for dance, and can thus be necessary prerequisites to acquire a certain component of the self-concept. They can also function as the representation of past social experiences to stabilize and maintain those aspects of self related to the past, for instance, a photo album or memorabilia. A third function of objects can be as symbols of the self, for instance, the content of one's bookshelf or china cabinet. Objects are thus able to provide or reflect social experiences and have to be considered just as others in the ecology of the self.

Environments or settings are also to be considered. They provide the place for a person's experiences and actions and can also be symbols of one's identity, as we have just heard from the work by Graumann and Schneider, or as we know from Proshansky (1978). Settings provide societal rules of conduct, as shown in Barker's (1965) concept of the behavior setting. Just as certain objects, so do they, too, provide the prerequisites for self-relevant behavior, such as a mountain for a skier or a lake for a sailor. But environments can also be arranged, for instance, inside living spaces. The arrangement and creation can be reflection and protection of the self and can be used also for self-presentational purposes (Kruse, 1980; Vinsel, Brown, Altman & Foss, 1980).

While none of these aspects has been neglected by psychological theorizing and research on the self, they have usually been considered separately. Taken together, however, the picture emerges of an ecological system. The self exists in interdependence with its ecology of others, things and environments. As long as the ecology of the self is stable, the self-concept will be stable and strive toward maintenance.

Self-concept change, however, results from an imbalance in the ecology of the self that leads toward restabilization under different ecological conditions, a restructuring of the ecology of the self.

This is the basic theoretical framework from which we consider the role of urban environments for a person's self-understanding, namely as one element of an eco-system (cf., Hormuth, 1984a). These elements can usually only be considered as part of the whole system. The present paper will try to assess the relative role of urban environments as compared to other aspects in the environment of a person considered by that person to be belonging to the self.

The method of assessment was the use of photographs. Similar approaches are known as autophotography from the work of Ziller and his associates (e.g., Ziller & Smith, 1977). Subjects are given use of a camera to take photographs of persons, things and environments which they consider to be part of or expressive of their self. One approach to explain this procedure to participants is to imagine an uncle in Australia. He does not know German, they do not know English; but they would like to show this uncle what kind of a person they are by way of these pictures. In our research, polaroid cameras are being used. Participants receive a film with ten pictures of which they should return seven. This allows them immediate control of the pictures, and retakes, if necessary. The content of each photograph is commented by the participant in one line.

The pictures are then rated by three independent raters. In a first step the content is being rated as belonging to either one of three categories: persons, things, and environments. All three can be represented within one photograph. In that case the relative percentage of the area covered by objects belonging to either one of the three is estimated. The next step of rating is more specific and, for instance, describes the kind of persons depicted (e.g., adults, children etc), or

the kinds of environments (e.g., nature, urban, etc.). Up to this step, the whole content of the picture was considered. In the next step, assessing the meaning of the picture, the rater has to decide which of the three categories is dominant in the picture. This decision is aided by the short, one line description of the content of the picture given by the research participant herself or himself. This main content is then rated on seven categories, based on our understanding of the function and meaning of persons, things, and environments for a person's self-concept.

The first category is the perspective taken, that is, a close-up versus very distant approach to the object depicted. For instance, research by Wofsey, Rierdan and Wapner (1979) has suggested this to us as an indicator of psychological closeness or involvement.

Then the integration of the object in the context of the picture was rated. Was the object presented in a more isolated way or integrated in a context? The next category tries to assess self-presentational aspects of the object photographed: objects that give a clear message about what kind of a person one is. Also assessed was the instrumental, activity encouraging aspect of the content of the picture. Is it, for instance, a tool for certain sports activities, like a tennis racket, or an environment, like a mountain for a skier?

If the picture and the related commentary indicated that the person had meant the content as an opportunity to reflect on some greater issue, for instance, gas pumps standing for the energy crisis, a high reflection score was given.

Some objects have value mainly for the memories that they revive, they are links to the past and thus can be used to create stability for one's self. For this aspect a memory value was scored. The next rating finally considered the symbolic value of the picture: Did the picture stand for something different than what it actually depicted, for instance, a computer for unemployment?

The currently available set consists of 488 photographs, taken by 91 subjects in the context of a larger project on relocation. This topic might conceivably make urban environments somewhat more salient (Hor-muth, 1984b). Due to the current status of the rating of the pictures, the data relating to meaning are based on 234 photographs on which rater agreement was satisfactory. Description of content refers to all 488 photographs. Of these, 44, that is 9%, depicted urban environments, which was one of 14 content categories into which the photographs were classified. Three of these referred to environments (urban, inside of living space, or nature), three to people, and eight to things. Let us now compare some of the differences in the meaning.

First we want to compare within the types of environments: Urban, nature, and inside living space. Urban environments were rated as having less self-presentational value and as being less instrumental for specific activities as compared to the average of all other categories. When comparing the perspective taken, urban environments are represented more distant and less involved (2.6), whereas nature is represented most immediate, from a perspective where the person is involved, and the inside of an apartment or house is photographed from an intermediate perspective (3.0;  $p$  less .001). Urban environments serve least for self-presentational purposes (1.5), inside living spaces most, as could be expected. Urban environments are also lowest in their activity value as well as their reflection value. In summary then, it seems that as compared to the two other categories of environments, urban environments consistently have the lowest scores as being functional for or expressive of the self-concept.

This holds true even when all other categories of objects depicted are taken into account as well. Perspective is most distant when urban environments or pets (that is, cats or dogs) are photographed, close-ups are used for sports and art objects. Urban environments are slightly

more integrated in their context than the average of the other categories, but not as well as pets or nature. The self-presentational value of urban environments is lowest (1.6), it is highest for cars and musical instruments (3.3 and 3.0). The symbolic value of urban environments is also very low (1.6), only persons are lower (adults 1.5). Urban environments and objects of art were considered to be least inviting for activities. The memory value of urban environments was lowest together with children and pets (1.0), however, the highest one for objects of art was only 1.7. Urban environments are also relatively low on the reflectional ratings (1.4), where books and objects of art are highest. Only sports objects and personal belongings, like a pair of glasses, are somewhat lower here. Again, even when compared to all possible objects depicted in the photographs, urban environments seem to be less functional to the self-concept than most other categories.

Next, we undertook a second level of analysis asking for what kind of person are urban environments more important or under what circumstances. Even if they do not seem to be as relevant to the self-concept as some other aspects of the "Umwelt" of the self is, that does not have to mean that there may not be special circumstances or special kinds of persons for whom they are more relevant. Two thirds of our subjects had not taken photographs of urban environments at all, 27% had one, 3% two, 3% three, and 1% five photographs of urban environments among the seven pictures descriptive of their self-concept. For further analyses, we compared the group of persons having any photograph of an urban environment to the group of persons having none. These groups did not differ on moving status, age, or sex. An interesting difference was found on a psychological scale measuring the extent to which people orient themselves on external social cues, the self-monitoring scale (Snyder, 1974). A person high on self-monitoring would, for instance, in a foreign country carefully watch how the natives use knife and fork, whereas a person low on self-monitoring would eat just the way she always does and

likes to. Persons having photographs of urban environments are significantly higher on the self-monitoring scale, that is, more externally oriented in looking for social cues (11.0 versus 9.3, p less .025).

Participants in this research project were also asked for significant life events in the past year. Two life events differentiated the groups with and without photographs of urban environments, even though only at a level of significance better than .10: The "urban environmentalists" indicated more frequently that they had to make a new beginning in life during the last twelve months (a euphemism for something going wrong) and they were more likely to have undergone some severe illness. Persons for whom urban environments are important to their self-concept are more likely to be retired or unemployed, less likely to be civil servants, workers or housewives, whereas they are equally represented among middle level white collar workers, students and apprentices. There is no relation to the number of times one has moved in his or her life or to age.

The information obtained from the present data is limited and probably incomplete. An increase in the number of photographs entered into the analysis, which will be possible after additional ratings can be done, will provide some more information based on additional statistical analyses that can then be undertaken, for instance, multivariate analyses. Those analyses may be able to support what, at this point, can only be a preliminary interpretation of the results as presented here.

Overall, when urban environments are seen in the context of all the possible aspects of the environment that may be of relevance to one's understanding and/or presentation of self, they seem to be of limited relevance. Only about one third of all participants in the research project considered urban environments as related to their self-concept at all, and of those, the large majority had only one such

photograph. Of all photographs, only about 9% were of urban environments.

Their meaning for the self-concept, as defined by our limited, but theoretically based, rating categories, was also consistently lower for urban environments than for most other categories. When urban environments are considered, they seemed to our raters to be less functional and meaningful for the self on several dimensions. However, the groups of people for whom urban environments do seem to be more expressive of their self-concept are in some ways different from others: They are more outside oriented in their search for social cues, they underwent a recent life-event that was likely negative and they are somewhat more likely not to be participating in the work life. These life situations are times when people need more outside orientation because their internal stability has been questioned as a system for guidance. This seems to be the kind of person and the kind of situation for whom urban environments become more relevant: the person who has to look at what is around him, who has to look for orientation in the outside world.

These findings also can be related to a model of self-concept change that makes use of the ecological approach to the self-concept (Hormuth, 1984a,b). According to this model, some life-events destabilize an element of the ecology of the self, for instance, the social network, and question the self-concept. This is supposed to lead to an outside orientation to find new aspects that can eventually be incorporated into a changed self-concept. An orientation to new aspects of social life could already be demonstrated in the life-style of persons who had moved to a new location in connection with a self-concept questioning life-event. The present data extent these findings by demonstrating an outside orientation toward the physical environment after some negative life events and connecting this to a psychological state of being receptive for outside cues in social behavior, namely the concept of self-monitoring.

## References

- Barker, R. G. (1965). Explorations in psychological ecology. American Psychologist, 20, 1-13.
- Cooley, C. H. (1902). Human nature and the social order. New York: Charles Scribner's Sons.
- Csikszentmihalyi, M., & Rochberg-Halton, M. (1981). The meaning of things: Domestic symbols and the self. New York: Cambridge University Press.
- Graumann, C.F. (1974). Psychology and the world of things. Journal of Phenomenological Psychology, 4, 389-404.
- Hormuth, S.E. (1984a). Restructuring the ecology of the self: A model for self-concept change. Paper presented at the International Conference on Self and Identity, Cardiff, Wales.
- Hormuth, S.E. (1984b) Transitions in commitments to roles and self-concept change: relocation as a paradigm. In V. Allen & E. van de Vliert (eds.), Role transitions: Explorations and explanations. New York: Plenum Press.
- James, W. (1890). The principles of psychology. New York: Holt.
- Kruse, L. (1980). Privatheit als Problem und Gegenstand der Psychologie. Bern: Huber.
- Mead, G. H. (1934). Mind, self, and society. Chicago: University of Chicago Press.
- Proshansky, H.M. (1978). The city and self-identity. Environment and Behavior, 10, 147-169.
- Shrauger, J.S., & Schoeneman, T.J. (1979). Symbolic interactionist view of the self-concept: Through the looking-glass darkly. Psychological Bulletin, 86, 549-573.
- Snyder, M. (1974) The self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 526-537.

Vinsel, A., Brown, B.B., Altman, I., & Foss, C. (1980). Privacy regulation, territorial displays, and effectiveness of individual functioning. Journal of Personality and Social Psychology, 39, 1104-1145.

Wofsey, E., Rierdan, J., & Wapner, S. (1979) Planning to move: Effects on representing the currently inhabited environment. Environment and Behavior, 11, 3-32.

Ziller, R. C., & Lewis, D. (1981). Orientations: Self, social, and environmental percepts through auto-photography. Personality and Social Psychology Bulletin, 7, 338-343.

Ziller, R.C., & Smith, D.E. (1977). A phenomenological utilization of photographs. Journal of Phenomenological Psychology, 7, 172-185.

Environments  
only

	Urban env.	inside living space	nature
perspective	2.6	3.0	3.9
self-presentation	1.5	1.9	1.7
activity	1.1	1.9	1.3
reflection	1.3	1.5	1.9

Environments Persons Objects	urban env.	nature	adults	children	pets	cars	musical instr.	objects of art	tools for sport
perspective	2.7				2.7			3.8	3.9
integration	3.0	3.3			4.0		1.8		1.4
selfpresentation	1.6					3.3	3.0		
symbolic value	1.6		1.5			3.1			
activity	1.3						3.2	1.2	4.4
memorial value	1.0				1.0	1.0			1.7
reflection	1.4							2.1	1.2

Appendix

Self-concept relevant ratings  
of picture content  
(1 - low 5- high)

Telefoninterviews in Sozialforschung und Sozialpsychologie:  
Ausgewählte Probleme der Stichprobengewinnung,  
Kontaktierung und Versuchsplanung

Stefan E. Hormuth & Erika Brückner  
Psychologisches Institut Max-Planck-Institut für  
Universität Heidelberg Bildungsforschung Berlin

Die hier berichteten empirischen Arbeiten entstammen zum großen Teil aus einem DFG-Projekt des ersten Autors (DFG Az: Ho 905/1-1f). Adressen der Autoren: Stefan Hormuth, Psychologisches Institut, Universität Heidelberg, Hauptstr. 47-51, D-6900 Heidelberg 1; Erika Brückner, Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, D-1000 Berlin 33.

### Zusammenfassung

Das Telefoninterview als Instrument sozialwissenschaftlicher Forschung beginnt sich auch in der Bundesrepublik zu etablieren, nachdem eine annähernd gesättigte Telefondichte erreicht ist. Da viele Erfahrungen nicht ohne weiteres aus den Vereinigten Staaten auf die deutschen Verhältnisse übertragbar sind, wird die Notwendigkeit einer eigenständigen Entwicklung und empirischen Überprüfung geeigneter Verfahrensweisen begründet. Randomisierte Generierung von Telefonnummern und Ziehung aus dem Telefonbuch werden empirisch miteinander in Hinblick auf die Ergebnisse von Kontaktversuchen und die Zusammensetzung der Stichproben verglichen. Die Anzahl notwendiger Kontaktversuche und günstige Zeiten werden beschrieben. Auf der Grundlage von Untersuchungen mit sozialpsychologischer Fragestellung wird die Realisierung experimenteller und quasi-experimenteller Versuchspläne verdeutlicht. Schließlich werden Besonderheiten der Datenqualität und der Konstruktion von Fragebögen erläutert.

### 1. Die Telefonumfrage als methodisches Dilemma

Das Telefon ermöglicht eine Form von Kommunikation, die als soziales Phänomen in der Forschung bislang keine Beachtung fand. Obwohl die rasche Verbreitung in den Industrienationen zu einem inzwischen selbstverständlichen Gebrauch dieses ältesten der neuen Medien führte, schenkte die Forschung dieser eigenständlichen Kommunikationsform keine Aufmerksamkeit. Weder als Forschungsgegenstand selbst noch als Forschungswerkzeug zur Kommunikation mit den zu untersuchenden Subjekten wurde das Medium in den Sozialwissenschaften thematisiert. Die Chancen eines leichten Zugangs zu breiten Bevölkerungsschichten in beliebiger geographischer Verteilung wurden in den USA relativ spät, im deutschen Sprachraum auch jetzt noch erst ansatzweise genutzt.

Wesentliche Einwände, das Medium zur Datenerhebung einzusetzen, wurden einerseits gegen die Repräsentativität von Stichproben vorgebracht, solange die Telefonversorgung nicht einen ausreichenden Prozentsatz der Haushalte umfaßte, andererseits aber gegen die eingeschränkten Möglichkeiten der telefonischen Kommunikation im Interview. Diese Bewertung drückt sich schon in der methodischen Nomenklatur aus. Dem 'fernmündlichen' Interview steht das mündliche gegenüber, das in der Fachsprache (exakter) als 'persönlich' deklariert ist, was unausgesprochen die telefonische 'Gesprächsform' als unpersönlicher klassifiziert.

Der Begriff der Umfrage ist in der deutschen Tradition fest gekoppelt mit dem Hausinterview und beinahe ein Synonym für persönliche Befragung. Die direkte Kommunikation zwischen Interviewer und Interviewtem gilt als der Königsweg der sozialwissenschaftlichen Umfrageforschung. Speziell dafür trainierte 'Datensammler' (selten die Forschenden selbst) begeben sich in die private Umgebung ihrer Informanten. Für die methodischen Vorzüge dieses Kommunikationsweges gibt es eine Reihe von Argumenten. Unter anderem wird die Sicherheit der Auswahl der Zielperson und die Beobachtungs- und Kontrollmöglichkeit durch den Interviewer beim Hausbesuch höher eingeschätzt als z.B. bei schriftlichen Interviews oder telefonischem Kontakt. Als überlegen gilt das persönliche Interview auch hinsichtlich der Interaktion in der Befragung.

Die Rapportmöglichkeiten des fernenmündlichen Kontakts erscheinen dagegen begrenzt wegen des Entzugs des optischen Kommunikationskanals. In den jahrzehntelangen Bemühungen, Datenerhebungen bei unterschiedlichsten Zielpopulationen exakt und standardisiert

durchführen zu können, haben sich für das persönliche Interview Umfragetechniken entwickelt, die auf der optischen Unterstützung der Übermittlung von Informationen beruhen. Die Notwendigkeit veränderter Operationalisierung von Fragestellungen und der Entwicklung neuer akustisch vermittelbarer Meßinstrumente haben die meisten Umfrageforscher bisher daran gehindert, den 'Königsweg' zu verlassen, nicht zuletzt, da dies eine Erschwerung der Replikation zuvor im persönlichen Interview durchgeföhrter Studien bedeutet.

In der Konsumentenforschung fanden Telefoninterviews in den USA schon frühe Verbreitung. Mit wachsender Telefondichte wurde das Verfahren dort auch schon in den vierziger Jahren zu Meinungsumfragen eingesetzt. Nicht zuletzt hatte aber die rasche Einföhrung eines methodisch noch wenig fundierten Verfahrens zu seinem schlechten Ruf beigetragen und die Anwendung in der sozialwissenschaftlichen Umfrageforschung eher verzögert. Telefonische Befragungen galten - nicht zuletzt auch aus Anlaß der falschen Wahlprognose eines Telefonsurveys - als "quick and dirty" und damit tabu für die seriöse Forschung. Es bedurfte einer langen Spanne der Methodenentwicklung und Rehabilitierung des Verfahrens, ehe die Telefonumfrage gegen Ende der siebziger Jahre in den USA, England und der Schweiz als gleichwertig neben postalischen und persönlichen Felderhebungen anerkannt wurde. Viele Sozialforscher begannen mit der Methode zu arbeiten und fundierten eine wissenschaftliche Auseinandersetzung, die mannigfaltige Veröffentlichungen hervorbrachte (u.a. Colombotos, 1969; Groves & Kahn, 1979; Herman, 1977; Jordan, Marcus, & Reeder, 1980; Miller & Cannell, 1962; Quinn, Gutek & Walsh, 1980; Rogers, 1976). Zusammenfassende Darstellungen der Methodendiskussion und -entwicklungen finden sich bei Dillman (1978) und Frey (1983).

Beachtung fand die Alternative nicht zuletzt auch deshalb, weil die Durchführung von persönlichen Interviews ebenfalls mit methodischen Problemen behaftet ist, die diesen 'Königsweg' als einziges aller möglichen Erhebungssettings in Frage stellen. Der Hausbesuch zum Interview ist nicht nur ein kostenaufwendiger Weg der Datenbeschaffung, sondern trifft auf zunehmende Schwierigkeiten. Die Realisierung von Interviews leidet unter Mobilität der Zielpersonen (Wohnungswechsel); erschwert ist auch der Zugang zu spezifischen Populationen und Untersuchungsgebieten (Erreichbarkeit von Berufstätigen, Abkapselung alleinstehender und älterer Frauen, Angst vor Kriminalität - auch seitens der Interviewer, generelles Mißtrauen gegenüber Fremden, schwerer Zugang zu Haushalten in bestimmten Siedlungsformen).

Spezifische Verzerrungen von Stichproben können die Folge sein und die methodische Qualität persönlicher Surveys möglicherweise ähnlich beeinträchtigen wie Ausfälle in ebenfalls spezifischen Teilstichproben (Haushalte ohne Anschluß) beim Telefonsurvey. Dies mag für amerikanische Verhältnisse stärkeres Gewicht haben und insbesondere auf urbane Gebiete zutreffen, kann aber auch bei deutschen Felduntersuchungen zunehmend zum Problem werden.

Repräsentative persönliche Befragungen stehen schon heute vor ähnlichen Problemen, wie sie für die USA geschildert wurden. Knapper werdende finanzielle Ressourcen für die Sozialforschung lassen methodisch hohe Ansprüche an persönliche Umfragen nur schwer durchsetzbar erscheinen. Steigende Fahrtspesen, Probleme der Erreichbarkeit von Zielpersonen und der Ablehnung an der Tür machen Hausbesuche teurer. Schulung und Kontrolle der Interviewer sind ebenfalls ein Kostenfaktor, der, verschärft durch Fluktuationen in den Interviewerstäben, nur schwer ohne Qualitätseinbuße kontrolliert werden kann.

Die Entwicklung von geeigneten, den Möglichkeiten des Mediums Telefon angepaßten Erhebungstechniken und ihre Anwendung in der amerikanischen wissenschaftlichen Umfrageforschung führte in den USA zu einer Legitimierung des Verfahrens. Bei einer dort über 90% liegenden, "gesättigten" Telefondichte und unter Verwendung eines speziellen Randomstichprobenverfahrens zur direkten Nummernauswahl (random digit dialing - RDD: Hauck & Cox, 1974; Klecka & Tuchfarber, 1978) können telefonische Erhebungen unter Aspekten der Repräsentativität den Sampling-Methoden des face-to-face Verfahrens überlegen sein.

Der Vorteil des telefonischen, preisgünstigeren, rascheren und gefahrlosen Zugangs zur Zielpopulation ist jedoch nur einer unter anderen. Die Tatsache, daß eine Umfrage von einem zentralen Studio aus durchgeführt werden kann, muß sich auf die Steuerung des Feldprozesses auswirken. Die zentrale Abwicklung von Erhebungen erlaubt eine direkte Supervision der Interviewer, die durch sofortige Überprüfung von Interviews und darauf folgende Rückmeldung die kontinuierliche Qualität der Befragung und eine erhöhte Interview- und Interviewerkonsistenz gewährleistet. Feedback und unmittelbarer Erfahrungsaustausch kann sich ebenfalls auf die Motivation der Interviewer und damit auf die Qualität ihrer Arbeit auswirken. Die mögliche Nähe des Forschers zu den durchführenden Organen und zum Gegenstand seiner Forschung im Feldprozess birgt auch die Chance neuer Erkenntnisse über Kommunikationsprobleme bei der Datenerhebung.

Darüber hinaus gestattet die räumliche Konzentration auch eine Nutzung technischer und elektronischer Hilfsmittel. Sukzessions- und Filterprobleme sowie Aufzeichnungsfehler werden durch Bildschirmvorgabe automatisch gesteuerter Fragenprogramme weitgehend ausgeschaltet. Das Telefon macht es möglich, die für Datendokumentation und Analyse inzwischen weit entwickelten elektronischen Techniken auch in den Erhebungsablauf so zu integrieren, daß eine Datenkontrolle noch während des Feldablaufs erfolgt. Die technische Perfektionierung, die neue Aspekte in die Datenerhebung einbrachte, sollte jedoch nicht darüber hinwegtäuschen, daß bis zu einer problemlosen Anwendung der Technologien noch Entwicklungsarbeiten geleistet werden müssen und weitere methodische Fragen zu klären sind. In der deutschen wissenschaftlichen Forschung stehen derzeit auch kaum technische Ressourcen und relativ wenig Erfahrungen mit wissenschaftlichen Telefonerhebungen zur Verfügung, so daß Forscher, die sich des Verfahrens bedienen wollen, nur begrenzt auf Routinen zurückgreifen können, sondern manche Verfahrensaspekte selbst entwickeln und erproben müssen.

Dabei kann die methodische Entwicklung und deren empirische Erprobung nicht allein auf bereits bestehenden Erfahrungen beruhen. Zum einen sind diese überwiegend in den USA erworben, so daß unterschiedliche Bedingungen der telefonischen Infrastruktur zu beachten sind. Zugleich aber sollte das Telefoninterview als Forschungsinstrument über den Einsatz im sozialwissenschaftlichen Survey hinaus weiterentwickelt werden und auf seine Eignung für spezielle Fragestellungen der Sozialpsychologie hin untersucht werden.

Die folgenden Ausführungen beschäftigen sich mit ausgewählten methodischen Fragestellungen des Telefoninterviews, die sich bei seinen Anwendungen im sozialpsychologischen Bereich stellten und über diesen engeren Bereich hinaus die generelle Verfahrensdiskussion berühren und praktische Problemlösungen aufzeigen. Im weiteren werden einige Beispiele die besonderen Möglichkeiten in der sozialpsychologischen Versuchsplanung demonstrieren.

## 2. Stichprobenziehung

Für das Telefoninterview ergeben sich gewisse Besonderheiten der Stichprobengewinnung, die einerseits - vor allem aufgrund des mangelnden persönlichen Kontaktes - Probleme schaffen, sich andererseits aber auch spezifische Merkmale des Telefonnetzes und seiner Infrastruktur zunutze machen können. Im folgenden soll auf einige dieser besonderen Probleme und Möglichkeiten eingegangen und Lösungsversuche auf der Grundlage

eigener Untersuchungen diskutiert werden.

Bei einer Telefonbefragung, die aus Anlaß der Ausstrahlung der Fernsehserie "Holocaust" 1978 in den USA und 1979 in der Bundesrepublik durchgeführt worden war (Hormuth & Stephan, 1981), wurde das Telefonbuch zur Auswahl der Stichprobe herangezogen unter Verwendung eines einfachen Zufallsschlüssels. In der Bundesrepublik 1979 noch existierende Bedenken gegen eine nicht ausreichende Telefondichte waren in dieser Untersuchung weniger wesentlich, da eine Überlappung der Telefonbesitzer und Fernsehapparatbesitzer und damit potentiellen "Holocaust"-Zuschauer angenommen wurde. Auf zufällig bestimmten Seiten wurde jeder zehnte Eintrag, der nicht eindeutig ein Geschäftsanschluß war, angerufen; falls kein Kontakt herstellbar war, der darauffolgende, und so weiter. Felddaten wurden unsystematisch erhoben, die Auschöpfungsquote wurde nach Befragung der Interviewer auf etwa 50% geschätzt. Während diese Methode und die Dokumentation ihrer Durchführung verbesserungswürdig war, so ist doch kennzeichnend, daß hier das Telefonbuch als Beschreibung der Grundgesamtheit verwendet wurde, d.h. die spezifische Infrastruktur des Telefonnetzes wurde genutzt. Dies ermöglichte eine zeit- und kostensparende Erstellung der Stichprobe, die den aktuellen Anlaß auch kurzfristig nutzen ließ.

Grundbedingung der Verwendung des Telefonbuchs als Beschreibung der Population ist eine Verbreitung des Telefons, die nur noch geringe Teile der Bevölkerung ausläßt. Repräsentative Bevölkerungsumfragen mit soziologischen und politologischen Fragestellungen scheiterten bis in die jüngste Zeit daran, daß nur ca. 80% aller Haushalte einen Telefonanschluß besaßen (ALLBUS, 1982). Inzwischen ist auch in der deutschen Forschung eine veränderte Situation eingetreten. Die Telefonversorgung stellt kein entscheidendes Hindernis mehr dar. Bei leichten regionalen Schwankungen und Abhängigkeiten vom Urbanitätsgrad nähern wir uns in der Bundesrepublik inzwischen (nach Analysen von Anders, 1982) einem Durchschittswert von über 90%. Da gewerbliche und private Anschlüsse nicht klar trennbar sind und Doppelanschlüsse eine exakte Zuordnung pro Haushalt nicht erlauben, kann 'Telefondichte' aufgrund offizieller Angaben nicht als exakter Richtwert gelten, sondern bedarf der empirischen Verifizierung.

Wenn man die Annahme einer annähernd gesättigten Telefonversorgung akzeptiert, so ist die folgende Frage die nach dem angemessensten Zugang zur Population und Auswahl der Stichprobe. Zwei verschiedene Modelle sind grundsätzlich abzuwägen, nämlich die Nutzung telefonspezifischer Information oder die weitgehende Übertragung bisheriger Methoden, wobei

dann erst nach der konventionellen Stichprobenziehung, z. B. über das Einwohnermeldeamt, für die entsprechenden Personen oder Haushalte die Telefonnummer festgestellt wird. Solche Methodenkontinuität erlaubt eine Vergleichbarkeit mit nichttelefonischen Stichproben, eine Bestimmung der nicht am Telefon erreichbaren Restgruppe und, wenn gewünscht, auch deren Befragung im persönlichen Interview.

Auf der anderen Seite steht für Telefonteilnehmer bereits eine öffentlich zugängliche Auflistung der Population zur Verfügung in Form des häufig zweimal jährlich erscheinenden Telefonbuches. Es gibt einige offensichtliche Unterschiede im Zugang zu der Stichprobe, z.B. in Bezug auf die Kosten. Wesentlicher allerdings ist die Datenqualität. Ohne empirische Vergleiche soll die Behauptung aufgestellt werden, daß normalerweise die Anmeldung eines Telefons vor der polizeilichen Anmeldung vorgenommen wird. Die Anmeldung beim Einwohnermeldeamt wird durch verschiedenste Faktoren beeinflußt, die wenig mit dem Wohnort zu tun haben. So unterbleibt z.B. eine Neuanmeldung, um noch im früheren Heimatort wählen zu können oder wählbar zu sein, oder um günstigere Versicherungstarife in Anspruch nehmen zu können. Die Anmeldung eines Telefons ist davon unbelastet. Sie gehört heute im Regelfall zu den ersten Dingen, die nach einem Zuzug als wesentlich erscheinen. Eine Eintragung in das Telefonbuch ist im Bereich der Deutschen Bundespost nur schwer zu unterdrücken, so daß die Telefonbücher tatsächlich eine im Vergleich zu den Informationen der Einwohnerbehörden nicht notwendigerweise schlechtere Grundlage zur Rekrutierung einer Stichprobe darstellen.

Selbstverständlich ist die Eignung dieser Informationsgrundlage auch von der jeweiligen Zielrichtung einer Untersuchung abhängig. Unter bestimmten Umständen sollen vielleicht gerade diese Restgruppen erfaßt werden, die im Telefonverzeichnis nicht aufgeführt sind. In solchen Fällen werden besondere Maßnahmen notwendig sein, so daß die Stichproben die Vor- und Nachteile verschiedener Quellen gegenseitig abwägend auf gemischter Datenbasis erstellt werden sollten.

Wenn zwischen der Beschreibung der Population im Telefonbuch und der tatsächlichen Population noch eine Diskrepanz besteht, so muß man sich anderer Methoden bedienen, um eine Stichprobe aus den tatsächlich existierenden Telefonanschlüssen zu ziehen. Solche Fehlerquellen, vor allem durch nichteingetragene Nummern und hohe Mobilität verursacht, haben in den USA zur Entwicklung des Random Digit Dialing, d.h. der Generierung von Telefonnummern aus zufälligen Zahlenkombinationen.

Ein Nichteintrag in das Telefonbuch ist in den Vereinigten Staaten

sehr einfach zu erreichen, nämlich im Regelfall durch eine Erklärung gegenüber der Telefongesellschaft und gegen eine geringe Gebühr. Die Furcht vor Belästigungen, vor allem in den Großstädten, führt dazu, daß diese Option oft wahrgenommen wird. Furcht vor Belästigung ist nicht gleichmäßig über die Population verteilt, sondern vor allem bei bestimmten Bevölkerungsgruppen zu finden. Ältere Personen, alleinstehende Frauen, oder Angehörige der im öffentlichen Bereich tätigen Berufe sind dabei weit überrepräsentiert. Dadurch ist ein amerikanisches Telefonbuch nur eine verzerrte Beschreibung der Population. Um dies auszugleichen, werden Stichproben aus der Grundgesamtheit aller möglichen Telefonnummern gezogen, d.h. es werden Zufallskombinationen gebildet, die angewählt werden. Wenn zwischen einer Telefonnummer und den Merkmalen ihres Inhabers kein Zusammenhang besteht, so stellt eine Randomstichprobe aus dem Universum der möglichen Telefonnummern auch eine Randomstichprobe aus allen Telefonteilnehmern dar. Durch den Einsatz technischer Hilfsmittel beim Wählen kann auch in Kauf genommen werden, daß für eine große Zahl der dabei zufällig hergestellten Telefonnummern kein Teilnehmer existiert.

Die Grundbedingungen, die zu einer hohen Zahl von "unlisted phone numbers" in den Vereinigten Staaten geführt haben, treffen im Bereich der Deutschen Bundespost nicht zu. Weder wird es dem Fernsprechteilnehmer einfach gemacht, seine Nummer zu verheimlichen, noch herrscht hier eine vergleichbare Furcht vor Belästigungen. Aus dem Fehlen dieser Grundannahmen ergibt sich, daß eine unkritische, d.h. empirisch nicht überprüfte, Übernahme dieser Methode aus den USA nicht sinnvoll ist. Dennoch kann es sein, daß bestimmte Bedingungen vorliegen (z. B. ein langer Zeitraum zwischen Anschluß und Eintrag, die private Verwendung von Geschäftsanschlüssen, oder die Nichtänderung von Einträgen, wenn der Telefonanschluß von anderen übernommen wird), die die Verwendung einer Zufallswählmethode anstelle des Telefonbuches als Grundlage nahelegen. Dabei ist allerdings zu bedenken, daß man sich bei ausschließlicher Verwendung von Zufallszahlen mancher Möglichkeiten der im Telefonbuch vorhandenen Information begibt, wo außer dem Namen auch z.B. Wohngebiet, Geschlecht, manchmal sogar Beruf und indirekte Hinweise auf die Nationalität zu finden sind.

Zur Überprüfung der Stichprobenzusammensetzung wurden empirisch zwei verschiedene Methoden der Stichprobengewinnung verglichen, nämlich (a) durch Zufallsziehung nach bestimmten Kriterien aus dem Telefonbuch gezogene Nummern, oder (b) die ebenfalls bestimmten einschränkenden Regeln unterworfen zufällige Generierung von Telefonnummern, eine

Modifizierung des Random Digit Dialing. Verglichen werden sollte die Zusammensetzung der Stichproben ebenso wie die Ergebnisse von Kontaktversuchen. Der Vergleich der beiden Methoden geschah im Rahmen einer telefonischen Befragung von 250 Personen im März 1983.

Die Telefonstichprobe wurde unter Verwendung des regulären Telefonbuches gewonnen, das zwei Monate zuvor erschienen war. Telefonnummern wurden nach dem Zufallsprinzip aus dem Stadtnetz Heidelberg gezogen. Solche Anschlüsse, bei denen es sich erkennbar um Geschäftanschlüsse handelte, wurden in die Stichprobe nicht aufgenommen. Ausländisch klingende Nachnamen wurden nicht ausgeschlossen, da diese zum einen kaum Rückschlüsse auf die Staatsangehörigkeit erlauben und in diesem Falle grundsätzlich alle der deutschen Sprache mächtigen Personen befragt werden sollten. Die Entscheidung, ob ein Interview mit einem Ausländer durchgeführt wurde, sollte erst nach einem Kontakt aufgrund der Sprachkenntnisse getroffen werden. Auf diese Weise wurden 312 Telefonnummern gezogen.

Die modifizierte Zufallsmethode ging davon aus, für jeden zufällig gezogenen Anschluß eine parallele Zufallsnummer zu generieren. Dabei sollte, unter Nutzung einer Eigenschaft von Telefonnummern, eine vergleichbare Streuung über den Stadtbereich erzielt werden, da sich Geschäfts- und Privatanschlüsse nicht zufällig über den gesamten Stadtgebiet verteilen. Eine Telefonnummer setzt sich aus zwei Teilen zusammen: einer zufälligen Zahlenkombination aus den vier letzten Ziffern der Nummer, und der Zahlenkombination des jeweiligen Bezirks, von der fünftletzten bis zur ersten Ziffer der Telefonnummer (je nach Größe der Stadt). Unter Beibehaltung der den Stadtbezirk kennzeichnenden Zahlenkombination wurden die letzten vier Ziffern des Telefonanschlusses permutiert. Dadurch wurden für jeden aus dem Telefonbuch gewonnenen Anschluß sechs entsprechende Nummern hergestellt (siehe Tabelle 1).

Tabelle 1 hier einfügen

Es ist notwendig, mehrere Parallelnummern herzustellen, da es sich bei der durch Permutation hergestellten Nummer entweder um einen Geschäftsanschluß oder eine nicht existierende Nummer handeln kann. Für diesen Fall diente dann die jeweils zweite oder dritte Permutation als Ersatznummer. Um einen Privatanschluß zu erreichen, genügte in 53% der Fälle die erste Permutation, in 42% mußte die zweite Permutation herangezogen werden, und nur in 5% war ein Rückgriff auf die dritte Permutation notwendig. Erkennbar ist, daß der Bedarf an Nummern sich hierbei im Gegensatz zur direkten Ziehung aus dem Telefonbuch etwa

verdoppelt.

Der erste empirische Vergleich beruht auf den Ergebnissen aller insgesamt 1436 Kontaktversuche, wobei für jeden Kontaktversuch ein Protokoll über Kontaktverlauf und -ergebnis geführt wurde. Für die Zufallszahlen handelt es sich um die Kontaktversuche mit den letztlich verwendeten Nummern. Wie aus Tabelle 2 ersichtlich, sind in den meisten Kategorien geringe Unterschiede festzustellen. Die Unterschiede werden deutlicher, wenn nur die Interviews berücksichtigt werden. 14.5% der Kontaktversuche mit generierten Nummern führen zu einem Interview, dagegen 19.5% der Kontaktversuche bei einer Nummer aus dem Telefonbuch. Dies ist allerdings nicht auf eine deutlich höhere Verweigerungsquote zurückzuführen, die auf Unterschiede in der Kontaktbereitschaft der erreichten Stichproben hinweisen würde. Etwas häufiger wurde nur ein Freizeichen gehört, hier handelt es sich um möglicherweise in den Abendstunden nicht identifizierte Geschäftsanschlüsse. Nach der Festlegung auf eine Nummer wurde nur noch in 1.2% aller Fälle ein Geschäftsanschluß erreicht, 0.8% mehr als bei der Auswahl aus dem Telefonbuch. Unterschiede in der Ausschöpfungsquote sind also auf die Summierung kleiner Unterschiede in den Stichproben zurückzuführen, die zu einer insgesamt etwas geringeren Präzision in der Stichprobe der Zufallsnummern führen, allerdings nicht zu schwerwiegenden Unterschieden.

#### Tabelle 2 hier einfügen

Damit ist aber noch keine Aussage gemacht über die tatsächliche Zusammensetzung der erreichten Personengruppen. Können durch die generierten Nummern Gruppen erreicht werden, die durch das Telefonbuch schwer auffindbar sind, z.B. Mobile, deren Nummer noch nicht eingetragen ist? Die tatsächlich interviewten Personen wurden in Hinblick auf mehrere demographischen Variablen verglichen. Das Ergebnis dieser Vergleiche auf insgesamt 19 Variablen ist summarisch in Tabelle 3 vorgestellt. Nur auf einer von 19 Variablen ist ein statistisch signifikanter Unterschied zwischen den beiden Gruppen festzustellen, was bei einem 5%-Signifikanzniveau nicht unerwartet ist.

#### Tabelle 3 hier einfügen

Auf der Grundlage dieser Daten scheinen die beiden Verfahren annähernd gleiche Populationen zu erfassen. Der etwas geringeren Präzision zufällig generierter Nummern steht der wesentlich höhere Arbeitsaufwand bei der Auswahl von Nummern aus dem Telefonbuch entgegen. Allerdings verlangt die hier vorgestellte modifizierte Methode eine vorherige Zufallsauswahl nach Stadtbezirken, am besten aus dem

Telefonbuch, aufgrund der für die Bestimmung der Population gewählten Kriterien, da sonst die Streuung der Zufallszahlen tatsächlich einen sehr großen Präzisionsverlust mit sich bringen kann. Nach der Ziehung einer kleinen Grundstichprobe kann dann eine wesentlich größere Stichprobe zufällig generiert werden. Wenn eine einfache Zufallsstichprobe benötigt wird, sollte es deswegen gegen diese Methode keine Einwände geben, da dadurch kurzfristig recht große Stichproben zusammengestellt werden können. Es besteht aber auch kein Grund zur Annahme, daß dadurch Populationen erreicht werden, die durch das Telefonbuch nicht erfaßt werden.

Diese Ausführungen versuchen nicht, die Qualität verschiedener herkömmlicher Methoden der Stichprobengewinnung, z. B. über das Einwohnermeldeamt, mit den hier vorgestellten Verfahren zu vergleichen. Vielmehr beschränkten sie sich auf einige Besonderheiten des Mediums Telefon und der Nutzung seiner Infrastruktur (unter den Bedingungen des deutschen Telefonnetzes). In diesem Rahmen gibt es Möglichkeiten der Stichprobengewinnung, die den Ansprüchen vieler Untersuchungsziele genügen und die den Stichproben, die aus anderen Auflistungen der Grundgesamtheit stammen, nicht notwendigerweise unterlegen scheinen.

### 3. Ergebnisse von Kontaktversuchen

Zu den Besonderheiten des Telefons, die es für Umfragezwecke sehr attraktiv machen, gehört die direkte, schnelle und preisgünstige Art der Kontaktierung im Gegensatz zum persönlichen Interview, das normalerweise einen oder mehrere Hausbesuche verlangt. Dadurch werden beliebige Wiederholungen von Kontaktversuchen ebenso wie solche zu verschiedenen Zeiten ermöglicht.

Als Kontaktversuch wird jeder Versuch bezeichnet, eine Telefonnummer aus der Liste der zufällig gezogenen oder generierten anzuwählen. Falls entweder ein Freizeichen zu hören oder keine Zielperson anwesend war, wurde die Nummer später, normalerweise zu anderer Tageszeit, wieder angewählt. Es stellt sich die Frage, wieviele Anwahlversuche sinnvoll oder notwendig sind, um eine Nummer endgültig zu bearbeiten. Eine steigende Anzahl von Kontaktversuchen verlangt wechselnde Tageszeiten und führt dadurch bald zu steil ansteigenden Kosten des einzelnen Kontaktversuchs, die sich dann noch erhöhen können, wenn schließlich andere Möglichkeiten der Kontaktierung eingesetzt werden sollen.

Abbildung 1, die auf 624 Kontaktversuchen basiert, zeigt den Prozentsatz der jeweils unbearbeiteten Nummern. Während nach dem ersten

Kontaktversuch bereits 45% der Nummern fertig bearbeitet waren, d.h. 55% verblieben, waren nach dem sechsten Kontaktversuch 95% aller Nummern tatsächlich kontaktiert, d.h. es kam entweder zum Interview, zur Ablehnung oder es konnte geklärt werden, daß es sich nicht um einen Zielhaushalt handelte. In Anbetracht einer Restgruppe von nur 5% scheinen mehr als 6 Kontaktversuche pro Nummer nur dann sinnvoll zu sein, wenn eine nahezu vollständige Aufklärung der Stichprobe angestrebt wird. Dies wird aber nicht mehr ausschließlich mit telefonischen Kontaktversuchen möglich sein.

Abbildung 1 hier einfügen

In diesem Zusammenhang sind die günstigsten Zeiten zur Kontaktierung einer Stichprobe von Interesse. Als günstig gelten normalerweise die frühen Abendstunden und die Nachmittage des Wochenendes. Vidgerhous (1981) gibt einige Daten an, die sich auf saisonale Unterschiede und Tagesschwankungen in Nordamerika beziehen. Solche Daten über bestimmte Stunden des Tages sind kaum auf die Bundesrepublik übertragbar, da z. B. das Fernsehen den Tag anders strukturiert. Die Interviews aus den hier berichteten Studien wurden üblicherweise zwischen 18 und 21 Uhr an Wochentagen durchgeführt, an Wochenenden von 11 (samstags) bzw. 12 (sonntags) bis 17 Uhr. Drei Variablen können dazu dienen, die günstigen Zeiten festzustellen: (a) die Freizeichen zu einem bestimmten Zeitpunkt, (b) der Prozentsatz der Interviews, der in einem bestimmten Zeitraum durchgeführt wurde, und (c) die Relation zwischen Interviews und Verweigerungen zu einem bestimmten Zeitpunkt. Diese letzte Zahl erlaubt eine Aussage darüber, zu welchen Zeiten Personen zwar anzutreffen sind, aber nicht gestört werden wollen. Die Tagesschau wird hierzu immer gern als Beispiel herangezogen.

Die folgende Abbildung zeigt für die drei Abendstunden von 18 bis 21 Uhr zuerst den Prozentsatz der Freizeichen, d.h. die erfolglosen Kontaktversuche. Hier ist die Zeit von 19 bis 20 Uhr die günstigste, nur etwa ein Drittel aller Anrufe wird nicht beantwortet, nach 20 Uhr ist nur noch bei jedem zweiten ein Kontakt möglich. Beim Verhältnis von Interviews zu Verweigerungen bedeutet die gestrichelte Linie ein Verhältnis von 1 zu 1, d. h. über der Linie überwiegen die Interviews, unter der Linie die Verweigerungen. Es zeigt sich hier, daß gerade die Zeit der besten Erreichbarkeit auch den höchsten Anteil an Verweigerungen erbringt, dagegen sind diejenigen, die zwischen 20 und 21 Uhr (d.h. während und nach der gefürchteten Tagesschau) erreichbar sind, auch am ehesten bereit, sich interviewen zu lassen. Wenn schließlich als dritte Variable der Anteil der in einem bestimmten Zeitraum

tatsächlich durchgeführten Interviews betrachtet wird, so schneidet die Stunde vor 21 Uhr auch hierbei recht gut ab.

Abbildung 2 hier einfügen

#### 4. Versuchsplanung

Ein Anspruch, der immer wieder vor allem an die Sozialpsychologie gestellt wurde, ist die Erweiterung der Basis der Merkmalsträger oder, in anderen Worten, die Einbeziehung anderer Populationen in die Datenerhebung als nur Studenten der ersten Semester. Dieser Anspruch, so selbstverständlich er erscheinen mag, ist nicht einfach zu erfüllen. Dabei handelt es sich um Probleme des Zugangs zu unterschiedlichen Populationen ebenso wie der dadurch entstehenden Kosten. Wenn auch in der deutschsprachigen sozialpsychologischen Literatur, etwas häufiger als in der amerikanischen, Untersuchungen zu finden sind, bei denen die Versuchspersonen Schüler der verschiedensten Schultypen oder sogar Arbeiter sind, so handelt es sich doch auch dabei um ausgewählte Populationen, die mehr der Einfachheit des Zugangs wegen ausgewählt wurden, als aus dem Bemühen heraus, die Merkmalsträgerbasis zu erweitern.

Die Methode des Telefoninterviews wird vor allem für die Umfrageforschung als geeignet angesehen, weniger aber in der sozialpsychologischen Forschung eingesetzt, wo eher die Herstellung experimenteller Bedingungsvariationen gefordert ist. Im folgenden soll anhand von empirischen Beispielen aus der Sozialpsychologie die Verwendung des Telefoninterviews als Instrument der Datenerhebung in experimentellen und quasi-experimentellen Versuchsplänen aufgezeigt werden.

Quasi-experimentelle Versuchspläne unterscheiden sich dadurch von Experimenten, daß die Zuordnung zu den experimentellen Bedingungen nicht zufällig vorgenommen werden kann (Cook & Campbell, 1979; Hormuth, Fitzgerald & Cook, 1985). Im Telefoninterview kann es sich dabei um Bedingungsvariationen handeln, die entweder aufgrund bekannter Information, d. h. meist durch den Telefonbucheintrag, erschlossen werden können, oder um solche, die erst im Laufe des Interviews erfragt werden können. Diese Unterscheidung hat Implikationen für den Versuchsplan und die Datenanalyse. Während im ersten Fall eine Steuerung bezüglich einer annähernden Gleichverteilung in verschiedenen experimentellen Bedingungen vorgenommen werden kann und damit eine annähernd gleiche Besetzung der Zellen, muß im zweiten Fall möglicherweise mit ungleichgewichtiger Zellbesetzung und ihren,

allerdings meist nicht unüberwindbaren Konsequenzen für die Datenanalyse gerechnet werden. Zwei Beispiele mögen dies illustrieren.

Im Rahmen der Überprüfung eines Modells zur Selbstkonzeptänderung sollten bestimmte Aspekte des Lebensstils und der Einstellungen von Personen nach einem Ortswechsel erfaßt werden (Hormuth, 1984). Dazu sollten Personen befragt werden, deren Umzug bis zu etwa einem Jahr zurücklag, sowie eine Kontrollgruppe aus Personen, die mindestens zwei Jahre an derselben Adresse wohnten. Diese Bedingungen für die Stichprobe konnten vollständig durch Nutzung der öffentlich zugänglichen Information aus den Fernsprechbüchern verschiedener Jahre erfüllt werden. Die Untersuchung wurde zuerst in den USA durchgeführt und dann in der Bundesrepublik repliziert. In den USA wurde das neue Telefonbuch, in der Bundesrepublik die durch die Post zur Verfügung gestellten Druckfahnen des neuen Telefonbuches auf nach dem Zufallsprinzip gewählten Seiten Eintrag für Eintrag verglichen, um so die im vergangenen Jahr hinzugekommenen Neueinträge identifizieren zu können. Durch die Verwendung der Telefonbücher vergangener Jahre konnte auch festgestellt werden, ob ein Anschluß über mehrere Jahre an derselben Adresse unverändert bestanden hatte. Auf diese Art wurde eine Stichprobe von 80% (400) Neuanschlüssen und 20% (100) Altanschlüssen zusammengestellt. Durch Befragung beim Interview ergab sich ein tatsächlicher Anteil von 70% (349) im letzten Jahr umgezogener Personen.

Allein durch die Verwendung des Telefonbuchs ließ sich so eine recht gute Präzision in der Zusammenstellung der Stichprobe erreichen. Es ergab sich weiterhin, daß durch diese Methode noch Personen erfaßt werden konnten, deren Umzug nur wenige Wochen zurücklag. Für die querschnittartige Beschreibung der sozialen Verhaltensweisen über ein Jahr nach einem Umzug wurde eine Gruppierung je nach dem Zeitpunkt vorgenommen. Von psychologischem Interesse im Rahmen des Modells der Selbstkonzeptänderung war, ob der Ortswechsel im Zusammenhang mit dem Beginn einer neuen sozialen Bindung oder dem Ende einer alten sozialen Bindung einherging, was durch Befragung der Interviewten erfaßt wurde. Die Häufigkeitsverteilungen in diesem faktoriellen Design sind in Tabelle 4 beschrieben.

#### Tabelle 4 hier einfügen

In einer anderen Untersuchung wurden die unabhängigen Variablen vorwiegend aufgrund im Interview erhobener Information bestimmt. Anlaß der Befragung war die Ausstrahlung der Fernsehserie "Holocaust", 1978 in den USA und im Januar 1979 in der Bundesrepublik. Überprüft werden sollte die sozialpsychologische Hypothese von der gerechten Welt (Lerner

& Miller, 1978), wonach unter bestimmten Umständen Opfern eigene Schuld an ihrem Schicksal attribuiert wird. Hormuth & Stephan (1981) versuchten, diese bisher überwiegend im Laborexperiment überprüfte Hypothese anhand der Darstellung der Judenverfolgung in ihrer Bedeutung für die Wirkung des Films zu testen. Verglichen wurden einerseits Zuschauer des Films mit solchen Personen, die den Film nicht gesehen hatten, andererseits wurde durch eine Serie von Fragen die Identifikation des Befragten entweder mit Opfern oder Machthabern in einer fiktiven, dem Holocaust vergleichbaren Situation erfaßt. Tabelle 5a zeigt den Versuchsplan, Tabelle 5b die tatsächliche Verteilung der Befragten über das faktorielle quasi-experimentelle Design.

Tabelle 5a und 5b hier einfügen

Zwei Probleme ergeben sich bei dieser Erfassung unabhängiger Variablen. Zum einen können bestimmte experimentelle Gruppen unterrepräsentiert sein. Tatsächlich identifizierten sich nur etwa 11% der Befragten mit den Machthabern, was aber bei über 250 Befragten eine sinnvolle statistische Regression bei ungleicher Zellbesetzung erlaubte. Das zweite Problem bezieht sich auf die kausale Richtung der Aussage: Besteht möglicherweise hier ein Zusammenhang zwischen dem Betrachten des Films und der Identifikation mit Opfern oder Machthabern? Der Argumentation kam hier zu Hilfe, daß der Anteil derjenigen, die sich mit den Machthabern identifizierten, unter den Zuschauern und Nichtzuschauern nahezu identisch war, nämlich 11.2% und 11.1%, was eine gegenseitige Beeinflussung dieser Variablen praktisch ausschließt.

Auf vergleichbare Weise machten sich Kommer & Schwarz (1984) das Telefoninterview bei der Untersuchung des Einflusses von Stimmung auf Urteilsprozesse zur Lebenszufriedenheit zunutze. Sie interviewten Personen unmittelbar nach einigen Spielen der Fußballweltmeisterschaft, die, je nach Ausgang, eine positive oder negative Stimmung hervorgerufen hatten. Ebenfalls im Bereich der kognitiven Sozialpsychologie setzte Tyler (1980) das Telefoninterview ein, um die kognitiven Prozesse bei der Beurteilung direkt und indirekt erfahrener Ereignisse am Beispiel der Kriminalität zu untersuchen. An diesen Beispielen zeigt sich auch, daß die Operationalisierungen theoretischer Gegebenheiten, die die Möglichkeiten des Telefons nutzen, an Unmittelbarkeit und Relevanz gewinnen können.

Die besondere Eignung für das sozialpsychologische Experiment und Quasi-Experiment ergibt sich aus den Möglichkeiten der Nutzung natürlicher Bedingungsvariationen und aus der Herstellbarkeit experimenteller Bedingungen durch die Variation von Stimulusmaterial und

Fragebögen, wodurch z.B. auch die Erforschung kognitiver Prozesse möglich ist. Selbst die Erfassung von Reaktionszeiten ist gut durchführbar. Beispiele der Überprüfung sozialpsychologischer Hypothesen und Theorien mit Hilfe des Telefoninterviews zeichnen sich durch eine breitere Streuung der Untersuchungsteilnehmer über verschiedene Altersgruppen und soziale Schichten aus. Durch die Möglichkeit und manchmal auch Notwendigkeit, natürlich vorkommende Ereignisse oder Bedingungsvariationen zu nutzen, gewinnen die Operationalisierungen an mundaner Realität im Gegensatz zu manchen Operationalisierungen, wie sie für Untersuchungen im Labor an Studenten der ersten Semester Psychologie vorgenommen werden.

##### 5. Datenqualität und Erhebungsmedium

Seit die telefonische Befragung auch in wissenschaftlichen Surveys in den USA eingesetzt wurde, steht das Problem der Validität und Qualität der mit dem neuen Medium erhobenen Daten im Mittelpunkt methodischer Überlegungen. Die Forschungen zur Überprüfung des Verfahrens beschäftigen sich deswegen hauptsächlich mit vergleichenden Studien von persönlichen und telefonischen Interviews. Obwohl der 'Stichprobenfehler' aufgrund der Unterschiede zwischen der Grundgesamtheit aller Haushalte und denen mit Telefonversorgung zunehmend an Bedeutung verlor, gaben die unterschiedlichen Übermittlungswege Anlaß zu Zweifeln an der Qualität der telefonisch erhobenen Daten. In Ergebnissen von Vergleichsstudien (Groves & Kahn, 1979; Quinn, Gutek & Walsh, 1980; Williams, 1977) zeigten sich jedoch kaum Abweichungen in Abhängigkeit vom Erhebungsmodus. Unterschiedliche Verteilungen bei Aggregaten von Antwortvorgaben traten zwar vereinzelt bei bestimmten Fragen auf, ließen jedoch keinen systematischen Erklärungszusammenhang mit dem Einfluß der Erhebungsmethode zu.

Ein Dilemma vergleichender Analysen ist die Unsicherheit hinsichtlich der 'wahren' Werte. Man mißt telefonisch erhobenes Datenmaterial an den im persönlichen Interview mit gleicher Fragestellung gewonnenen Informationen, die als Maßstab gelten, obwohl auch Fehlerquellen der persönlichen Befragung bekannt sind. Systematische Verzerrungen, die z.B. durch Interviewereinflüsse entstehen können, lassen sich im mündlichen/persönlichen Survey nie völlig eliminieren und noch schwerer erkennen oder hinsichtlich spezifischer Effekte exakt einschätzen. Ein in der natürlichen Umgebung des Befragten persönlich durchgeführtes Interview vermag zwar ein genaueres Bild über den Hintergrund des Datenmaterials liefern, da

falsche Angaben durch Augenschein erkennbar sind (z.B. Kerndaten über Familienstruktur, Wohnsituation, eventuell auch sozialen Status). Zur Datenkontrolle wird diese Möglichkeit de facto wenig eingesetzt. Sie liefert vielleicht auch nur wenig brauchbares Material, da Beobachtungen des Interviewers meist keine zuverlässige Datenquelle darstellen können. Die Vorteile der persönlichen Erhebungssituation sind - insbesondere bei Einsatz größerer Interviewerstäbe - schwer zu nutzen, während das Risiko durch persönliche Einflüsse (Interviewer-Bias) und unkontrollierbare technische Mängel in der Durchführung (z.B. Fehler in Sukzession, unexakte Vorgaben von Fragen und Beantwortungshilfen) ein belastender Faktor des Verfahrens ist, wie Untersuchungen zur Interviewerqualität häufig zeigen (Prüfer & Rexroth, 1984).

Ohne zu postulieren, daß Telefoninterviews aufgrund der reduzierten Einflüsse durch die Person des Interviewers und der zentral besser steuer- und kontrollierbaren technischen Qualität des Befragungsprozesses dem Hausbesuch überlegen sind, kann man einige unterschiedliche methodische Implikationen der beiden Befragungstechniken bereits erkennen. So nutzvoll vergleichende Datenanalysen zur Absicherung der oben beschriebenen Ergebnisse sein können, so sehr vernachlässigen sie die Tatsache der Eigenständigkeit beider Verfahren, deren spezifische Charakteristika als erklärende Variablen bisher zu wenig beachtet wurden. Für die fundierte Weiterentwicklung des Verfahrens sind aber noch weitere Untersuchungen notwendig, die spezifische Möglichkeiten und Probleme der auditiven Vermittlung nicht nur unter dem Blickwinkel des Vergleiches 'persönlich-telefonisch' systematisch erforschen.

Unter dem Aspekt der Datenqualität lassen sich auf dem Erfahrungshintergrund der bisher von den Verfassern durchgeföhrten Studien einige Hypothesen ableiten, die sich auf spezifische Charakteristika des Telefoninterviews stützen. Die Kommunikation zwischen Interviewer und befragter Zielperson ist nicht nur durch das Fehlen des Blickkontaktees von einigen zwar rapportfördernden, möglicherweise aber auch störenden Stimuli befreit, sondern bindet auch den Befragten direkter in den Interaktionsprozeß ein.

Der Befragte muß sich am Telefon mehr als im persönlichen Interview auf seinen unsichtbaren Gesprächspartner konzentrieren. Störungen durch Geräuschkulissen, insbesondere durch die Einmischung anderer anwesender Personen, versuchen die Befragten oft spontan selbst abzustellen. Dritte Personen, die anfangs meist neugierig das Interview verfolgen, verlieren häufig schon mangels Verständnis der Fragen ihr Interesse und

werden nicht selten ausgeschaltet, weil es dem Befragten lästig ist, die Kommunikation zu Dritten herzustellen. Das Telefoninterview stiftet also möglicherweise eine besonders enge Dyade zwischen den Interaktionspartnern, trotz eingeschränkter nonverbaler Stimuli. Die Ablenkungsmöglichkeiten sind geringer, was eine gute Voraussetzung bietet, sich voll auf den Interviewprozeß zu konzentrieren. Die Intimität der Gesprächssituation, die Dritte weitgehend 'draußen läßt', und eine fast ausschließliche und damit auch spezifische persönliche Verbindung zum Interviewer schafft - bei gleichzeitiger Distanz zur immerhin fernen, relativ anonymen Person des Interviewers - kann einen bedeutsamen Einfluß auf die Aussagebereitschaft haben.

Da der Befragte dem Interviewer keine Einsicht (= optisch/soziale Kontrolle) in seine Privatsphäre geben muß, kann er sich freier fühlen. Diese Bedingungen können zu ehrlicheren Antworten führen, wie freimütige Äußerungen von Interviewpartnern zeigten. Ein systematischer Einfluß solcher Phänomene auf die Datenqualität konnte bisher nicht nachgewiesen werden. Andrerseits stellten Untersuchungen von Groves und Kahn (1979) und Jordan (1980) fest, daß bei Telefonsurveys höhere Raten an "weiß nicht"-Angaben und Verweigerungen auf heikle Fragen als bei persönlichen und schriftlichen Befragungen auftraten. In Fällen besonders sensitiver Fragen können auch am Telefon Verfahren zur Sicherung der Vertraulichkeit der Information, z.B. randomisierte Antworttechniken, eingesetzt werden (Hormuth & Boruch, im Druck; Orwin & Boruch, 1982).

Eine telefonische Umfrage an zufällig ausgewählten 250 Personen (Hormuth, 1983), ermöglichte einen Vergleich psychologischer Instrumente, die entweder postalisch versandt und schriftlich oder bei einem Hausbesuch persönlich erläutert und ebenfalls schriftlich beantwortet oder aber am Telefon erhoben wurden. Dies kann am Beispiel einer Skala verdeutlicht werden, die Zufriedenheit mit dem Selbst und Änderungsbereitschaft erfassen soll.

Die Skala wurde für den Einsatz am Telefon nur geringfügig überarbeitet, indem einige wenige Items gestrichen wurden, vor allem aber, indem die Antwortskala von einer zwar optisch gut repräsentierbaren, aber kognitiv weniger gut verankerten 7-Punkt-Skala in eine 5-Punkt-Skala umgewandelt wurde. Die Skala, die durchaus intime Fragen beinhaltet, erzielte auch in der verkürzten Form bei einer heterogenen Zufallsstichprobe in der telefonischen im Vergleich zur mehr homogenen Stichprobe in der schriftlichen Beantwortung vergleichbar gute Reliabilitäten für die Gesamtskala und die Unterskalen; es gab keinerlei Verweigerungen der Antworten an diesem Punkt der Befragung. Auch

psychologische Tests konnten nach einer stilistischen Überarbeitung und Modifizierung der Antwortkategorien am Telefon problemlos eingesetzt werden.

Tabelle 6 hier einfügen

Ein weiteres, vielleicht noch zentrales Problem im Zusammenhang mit der Frage nach der Datenqualität ist Exaktheit und Komplexität der Informationsvermittlung. Nach Kriterien der Plausibilität wird angenommen, daß eine Beschränkung auf den auditiven Kanal das kognitive Niveau entscheidend reduziert. Fragetexte und Itemskalen müssen demnach im Volumen und ihrer Verständlichkeit stark vereinfacht werden, weil vorausgesetzt wird, die Befragten können nur simple, kurze Sätze und Antwortkategorien verstehen und behalten. Berechtigt ist diese Annahme, wenn man sie auf Befragungstechniken aus der persönlichen Umfrageforschung bezieht, in denen längere Statements, die oft sogar mehrere Stimuli enthalten, von Befragten (notfalls mehrmals) gelesen werden können oder Listen mit Antwortvorgaben die auch nur akustisch gestellte Frage kognitiv unterstützen. Visualisierungen in Form von Feldern, Leitern oder Thermometern sind sicherlich nützliche Hilfen für den Befragten. Entscheidungsprozesse, die eine Reihe möglicherweise sehr unterschiedlicher Alternativen einbeziehen, sind nur mit Hilfe von Karten möglich, damit der Befragte sich einen Überblick verschaffen und Auswahlprozeduren durchführen kann.

Einige, bei persönlicher Befragung häufig eingesetzte Operationalisierungen, wie Rangordnung oder Sortieren von Items sind nicht telefonisch replizierbar, da sie auf der Basis optischer Befragungshilfen konzipiert sind und ohne diese sinnlos werden. Diese Tatsache und der Versuch, solche Operationalisierungen an die veränderten kognitiven Bedingungen anzupassen, verführt leicht zu der Annahme, daß der akustische Vermittlungsprozeß prinzipiell Defizite hat und keine wissenschaftlich vertretbaren Meßtechniken zuläßt. Daraus resultierende Unsicherheiten führen häufig zu einer übertriebenen Vereinfachung bei der Fragebogenkonstruktion. Man beschränkt sich auf relativ kurze, möglichst dichotome Fragen und bevorzugt "Ja/Nein"-Kategorien. Verbale Bewertungs- und Frequenzskalen, die häufig zwischen vier und sechs Ausprägungen umfassen, können zwar vermutlich telefonisch noch gut verstanden und behalten werden, wenn die einzelnen Kategorien kurz sind, ihre Eignung wird jedoch angezweifelt. Während die optische Vorlage solcher Skalen dem Befragten ein klar strukturiertes Gerüst gibt, an das er sich bei seiner Differenzierung halten muß (normative Wirkung), werden die Antwortkategorien bei akustischer

Aufnahme unter Umständen nur teilweise verstanden und möglicherweise vom Befragten unter Verwendung anderer, eigener Adjektive modifiziert. Einige der bekannten kategorialen Skalen sind erkennbar methodisch für Telefoninterviews ungeeignet, da sich die Items semantisch als unprägnant erweisen für diese akustische Übermittlungsform (z.B. 'sehr stark', 'eher stark').

Die Einsetzbarkeit von numerischen, mehrstufigen Skalen mit zwei bis drei verbalisierten Verankerungen (Anfangs-, End- und möglicherweise Mittelpunkt) wurde in der Methodendiskussion kaum erörtert oder auch von vornherein bestritten, da die häufig im persönlichen Interview gebrauchten Zahlenreihen (0 oder 1 bis 7, 10, usw.) durch gleichgroße Felder oder Linealmarkierungen visualisiert werden, deren Endpunkte dann deklariert sind. Die Verwendbarkeit von Skalen am Telefon verlangt eine deutliche kognitive Repräsentation der Endpunkte, z.B. Null und zehn, während die Zwischendifferenzierungen nach Möglichkeit zur Vermeidung übermäßiger Komplexität nicht verbal vorgenommen werden sollten.

In Studien von Hormuth und Stephan (1981) wurden Skalen von Eins bis Neun von diesen Autoren mit gutem Erfolg telefonisch implementiert, später dann auch Zahlenräume von Eins bis Zehn (vgl. Brückner, Hormuth & Sagawe, 1982). Es zeigte sich, daß die Befragten mit den vorgegebenen Werten nach nur kurzer Einübung umgehen können. Der Zahlenraum ist auch ohne optische Hilfe in der Vorstellung reproduzierbar und zwei prägnante Bezeichnungen der Endpunkte lassen sich leichter erfassen als mehrere meist sehr ähnliche Items, vor allem bei Verbalskalen. Ein Wechsel der Inhalte bzw. der Extrempunktkategorien erwies sich als problemlos. Da Skalen dieser Art viele Operationalisierungsmöglichkeiten bieten, kann ein Befragungsinstrument für Telefoninterviews durchaus differenziert gestaltet werden. Eine weitgehende Beibehaltung der Antwortskala ermöglicht eine Konzentration auf die Inhalte und Items des Erhebungsinstruments, die durch ihre Gestaltung und Abwechslung einer befürchteten Monotonie entgegenwirken können.

Ein Teil der methodischen, instrumentellen Probleme, die sich bei geplantem Methodenmix und Replikation von mündlichen/persönlichen Umfragen stellen, lässt sich somit lösen. Vergleichbare Datenqualität und Übertragbarkeit von Meßinstrumenten sollte jedoch nicht vergessen lassen, daß der Erhebungsprozeß selbst sehr unterschiedlich ist. Die notwendige Anpassung von Meßinstrumenten aus der persönlichen Befragung auf die telefonische kann nicht nur durch einfaches Weglassen optischer Hilfsmittel realisiert werden. Bei paralleler Implementierung beider Verfahren sollte man konsequenterweise generell auf solche

Befragungshilfen verzichten, was einen Bruch mit bewährten Traditionen bedeuten würde. Die Verfasser vertreten deshalb die Auffassung, daß dann Operationalisierungen primär für die telefonische Erhebung konzipiert werden sollen, was ihre problemlose Anwendung auch im persönlichen Interview zuläßt. Mit der wissenschaftlichen Legitimierung der telefonischen Datensammlung und einer Intensivierung der Methodenentwicklung im Fragebogenbereich werden sich Transferprobleme, die derzeit eine Hürde für manchen Forscher darstellen, weiter abbauen.

Die bisherige sozialwissenschaftliche Verwendung der Methode beschränkte sich überwiegend auf Übertragungen quantitativer Befragungen (insbesondere von politischen Einstellungen). Eine breite Erfahrungsbasis gibt es in der Bundesrepublik erst im Bereich von Marktuntersuchungen. Die Verwendbarkeit psychologischer Tests am Telefon wurde häufig bezweifelt, ist aber für den sinnvollen Einsatz der Methode innerhalb der Psychologie notwendig. Der Einsatz des Telefoninterviews in der Sozialpsychologie und Nachbargebieten hängt auch davon ab, daß spezifisch psychologische Inhalte erhebbar sind.

Diese Erfahrungen haben gezeigt, daß das Telefoninterview sinnvoll das sozialwissenschaftliche Methodenrepertoire ergänzen kann. Für die Methode sprechen auch die niedrigeren Kosten und der vergleichsweise geringe Zeitaufwand bei der Datenerhebung. Bei eigenen Untersuchungen konnten typischerweise etwa 250 Interviews in neun Tagen, d.h. an zwei Wochenenden und in der dazwischenliegenden Woche, mit fünf bis sieben Interviewern realisiert werden. Die Telefondichte und die relativ zuverlässigen und oft erneuerten Informationen über die Population in öffentlich zugänglichen Telefonbüchern erleichtern die Bestimmung der Stichprobe.

Darüber dürfen Nachteile des Telefons als Erhebungsinstrument nicht vergessen werden. Die Kommunikation ist auf den auditiven Kanal beschränkt. Streng kontrollierte experimentelle Manipulationen sind kaum durchzuführen, nicht zuletzt, weil ihr Erfolg und eine notwendige anschließende Aufklärung schwer überprüfbar sind. Telefonische Verfahren sollten dennoch nicht nur unter den Aspekten der substituierenden, nur erhebungsökonomischen Alternative zum 'Königsweg' der empirischen Sozialforschung angesehen werden, sondern als eine eigenständige Entwicklung, die methodische Investitionen und empirische Überprüfungen erfordert und ebenso wie alle anderen Methoden sozialwissenschaftlicher Forschung mit den ihr eigenen Vorteilen und Nachteilen am ehesten in Kombination mit verschiedenen Erhebungsverfahren bestehen kann.

## Literatur

ALLBUS (1982) Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (1982). Codebuch, Methodenbericht und Vergleichsdaten. Köln: Zentralarchiv für empirische Sozialforschung.

Anders, M. (1982). Das Telefoninterview in der Bevölkerung: Voraussetzungen, Verfahren, Vorteile. München: Infratest.

Brückner, E., Hormuth, S.E., & Sagawe, H. (1982). Telefon-interviews: Ein alternatives Erhebungsverfahren? ZUMA-Nachrichten, 11, 9-36.

Colombotos, J. (1969). Personal versus telephone interviews: Effect on responses. Public Health Reports, 84, 773-782.

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design and analysis issues for field settings. Chicago: Rand McNally.

Dillman, D. A. (1978). Mail and telephone surveys: The total design method. New York: Wiley.

Frey, J. H. (1983). Survey research by telephone. Beverly Hills, CA: Sage.

Groves, R. M. & Kahn, R. L. (1979). Surveys by telephone: A national comparison with personal interviews. New York: Academic Press.

Hauck, M., & Cox, M. (1974). Locating a sample by random digit dialing. Public Opinion Quarterly, 38, 253-260.

Hormuth, S.E. (1983). Ortswechsel als Gelegenheit zur Änderung des Selbst. Bericht und Fortsetzungsantrag an die DFG. Psychologisches Institut der Universität Heidelberg.

Hormuth, S.E. (1984). Transitions in commitments to roles and self-concept change: Relocation as a paradigm. In V. Allen & E. van de Vliert (Eds.), Role transitions: Explorations and explanations. New York: Plenum Press.

Hormuth, S.E., & Stephan, W.G. (1981). Effects of viewing "Holocaust" on Germans and Americans: A just-world analysis. Journal of Applied Social Psychology, 11, 240-251.

Hormuth, S. E., & Boruch, R. F. (im Druck). Schutz der Privatsphäre wider wissenschaftliche Neugierde: Personen- und informationsbezogene Verfahren zur Lösung des Interessenkonflikts. In A. Spitznagel & L. Schmidt-Atzert (Hg.), Sprechen und Schweigen: Zur Sozialpsychologie der Selbstenthüllung. Bern: Huber.

Hormuth, S. E., Fitzgerald, N. M., & Cook, T. D. (1985). Quasi-experimental methods for community based research. In E. C. Susskind & D. C. Klein (Eds.), Community psychology: Methods, paradigms, and applications. New York: Praeger.

Jordan, L. A., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. Public Opinion Quarterly, 44, 210-222.

Kegeles, S. S., Fink, C. F., & Kirscht, J. P. (1969). Interviewing a national sample by long-distance telephone. Public Opinion Quarterly, 33, 412-419.

Klecka, W. R., & Tuchfarber, A. J. (1978). Random digit dialing: A comparison to personal surveys. Public Opinion Quarterly, 42, 105-114.

Kommer, D., & Schwarz, N. (1984). Panem et circenses: Effects of irrelevant events on judgments of well-being. Manuskript: Psychologisches Institut der Universität Heidelberg.

Orwin, R. G., & Boruch, R. F. (1982). RTT meets RDD: Statistical strategies for assuring response privacy in telephone surveys. Public Opinion Quarterly, 46, 560-571.

Prüfer, P. & Rexroth, M. (1984) Beschreibung eines Verfahrens zur Bewertung von Interviewerverhalten. ZUMA-Arbeitsbericht, IV/84. Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Quinn, R. P., Gutek, B. A., & Walsh, J. T. (1980). Telephone interviewing: A reappraisal and a field experiment. Basic and Applied Social Psychology, 1, 127-153.

Rogers, T. F. (1976). Interviews by telephone and in person: Quality of responses and field performance. Public Opinion Quarterly, 40, 51-65.

Schuman, H. & Presser, S. Question answers in attitude surveys: Experiments on question form, wording, and context. New York: Academic Press.

Sudman, S. (1977). New uses of telephone methods in survey research. Journal of Marketing Research, 3, 163-167.

Tuchfarber, A. & Klecka, W. (1976). Random digit dialing: Lowering the cost of victimization surveys. Washington, DC: Police Foundation.

Tyler, T.R. (1980). Impact of directly and indirectly experienced events: The origin of crime related judgments and behaviors. Journal of Personality and Social Psychology, 39, 13-28.

Vidgerhous, G. (1981). Scheduling telephone interviews: A study of seasonal patterns. Public Opinion Quarterly, 45, 250-259.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication. Psychological Bulletin, 84, 963-976.

**Tabelle 1:****Verfahren der modifizierten Zufallsgenerierung von Telefonnummern****1. Schritt: Zufallsauswahl aus dem Telefonbuch**

65 4321

Bezirk      Permutierbare Nummer

**2. Schritt: Permutationen**

	Fünfstellig	Sechsstellig
Original:	5 4321	65 4321
Permutationen:	5 4312	65 4312
	5 3421	65 3421
	5 4231	65 4231
	usw.	usw.

**3. Schritt: Anwahl**

Falls Freizeichen:

zu anderer Tageszeit anrufen

Falls "Kein Anschluß":

nächste Permutation

Falls "Kein Privatanschluß":

nächste Permutation wenn nicht bereits  
sechster Kontaktversuch

Tabelle 2:

Ergebnisse der Kontaktversuche bei zufällig generierten Nummern und Anschläßen aus dem Telefonbuch

	Zufalls- generierte (N=730)	Telefon- buch (N=706)
Interview	14.5%	19.5%
Keine Zielperson	1 %	0.5%
Termin vereinbart	11 %	12 %
Verweigert	19 %	18 %
Kein Privatanschluß	1.2%	0.4%
Freizeichen	50 %	48 %
Sonstige	3.3%	1.6%

Die Daten beruhen auf 1436 Kontaktversuchen Ende März 1983.

Tabelle 3:

Signifikanz der statistischen Tests der Unterschiede zwischen Merkmalen der aus dem Telefonbuch gezogenen und der durch Permutation hergestellten Stichproben.

Variable	Unterschied
Alter	n.s.
Alter vom Interviewer geschätzt	n.s.
Geschlecht der Interviewten	n.s.
Familienstand	n.s.
Schulbildung	n.s.
Haushaltsgröße	n.s.
Art der Wohnung	n.s.
Erwerbstätigkeit	n.s.
Beruf	n.s.
Berufl. Position	n.s.
Umzug im letzten Jahr	n.s.
Zeitpunkt des letzten Umzugs	n.s.
Dauer des Wohnverhältnisses	n.s.
Alleinlebend	n.s.
Wie lange nicht mehr tätig?	sig.
Geschlecht der Verweigerer	n.s.
Alter der Verweigerer (geschätzt)	n.s.
Geschlecht der Interviewer	n.s.
Interviewdauer	n.s.

Unterschied

Die Daten beruhen auf 246 Interviews im März 1983. Die Korrelation zwischen dem geschätzten Alter und den Angaben der Befragten betrug bei den Interviews  $r = .80$ , so daß diese Schätzungen auch bei den Verweigerern Hinweise auf das tatsächliche Alter geben können.

Tabelle 4:

**Verteilung der Neuzuzieher (N=346)**  
 (durch Befragung erfaßt)

**Zeitraum seit Umzug in Monaten**

	0 - 1	2 - 3	4 - 6	7 - 9	10-12	12+
n	17	32	51	84	89	73
%	4.9	9.2	14.7	24.3	25.7	21.1

**Umzugsanlaß**  
 (Selbst-  
 einschätzung)

Ende einer Bindung	182
	52.6
Beginn einer Bindung	124
	35.8
Sonstige	40
	11.6

Daten nach Hormuth (1984).

Tabelle 5a:

## Versuchsplan der Befragung zur Fernsehserie "Holocaust"

	Identifikation mit	
	Opfern	Machthabern
Nicht-	:	:
Zuschauer	:	:
	:	:
	-----	-----
	:	:
	:	:
Zuschauer	:	:
	-----	-----

Identifikation wurde erfragt:

"Versuchen Sie einmal, sich vorzustellen, daß sich ähnliche Ereignisse wie die des Holocaust wieder ereignen würden. Wenn Sie sich das vorstellen, sehen Sie sich selbst dabei auf der Seite der Opfer oder auf der Seite der Machthaber?"

nach Hormuth & Stephan (1981)

Tabelle 5b:

Zellverteilung bei der Befragung zur Fernsehserie "Holocaust"

## Identifikation mit

## Opfern

## Machthabern

	n = 96	n = 12
Nichtzuschauer	Gesamt: 36.9%	Gesamt: 4.6%
	Zeile: 89.9%	Zeile: 11.1%
Zuschauer	n = 135	n = 17
	Gesamt: 51.9%	Gesamt: 6.5%
	Zeile: 89.8%	Zeile: 11.2%

nach Hormuth &amp; Stephan (1981).

Tabelle 6:

Reliabilitäten der Selbständerungsskala in verschiedenen

Untersuchungen

(beruhend auf Cronbachs alpha)

	Befragungsart		
	Postalisch	Persönlich	Telefonisch
<b>Unterskala</b>			
Dinge	.73	.75	.48
Familie	.78	.77	.64
Arbeit	.87	.80	.85
Sozial	.83	.67	.61
Global	.86	.82	.72
<hr/>			
Änderungs-			
bereitschaft	.84	.76	.68
Zufriedenheit	.85	.81	.72
<hr/>			
Gesamtskala	.91	.88	.80
<hr/>			
n	90	105	255
<hr/>			

Die Stichproben der drei Untersuchungen unterschieden sich in ihrer Homogenität. Nur die Telefonstichprobe war eine reine Zufallstichprobe und deswegen am heterogensten. Die beim Telefoninterview verwendete Form der Skala war von 30 auf 25 Items gekürzt. Eine Beschreibung der Skala findet sich bei Hormuth (1983).

Abbildung 1:

Anteil der bearbeiteten Telefonnummern pro Kontaktversuch

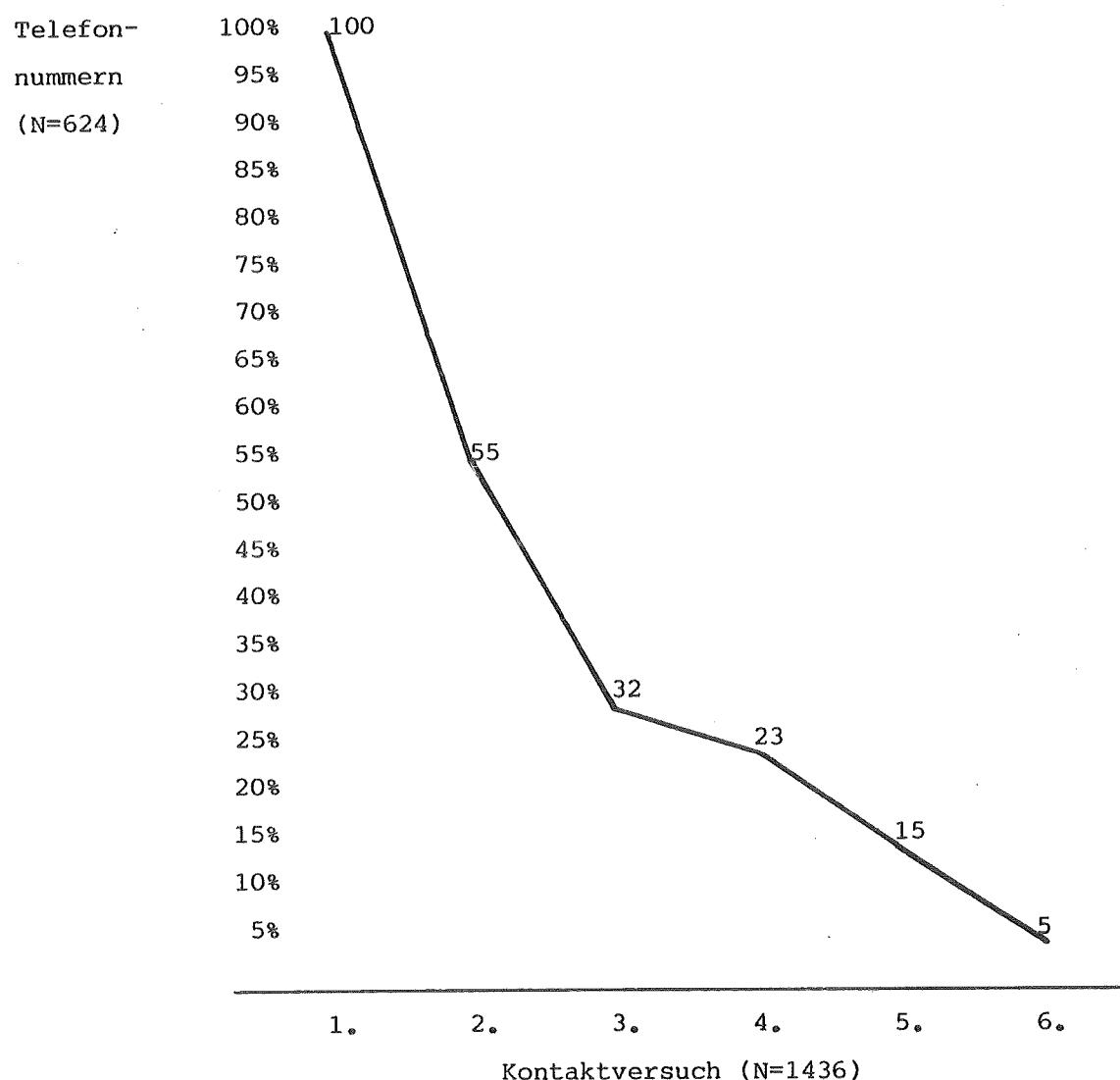
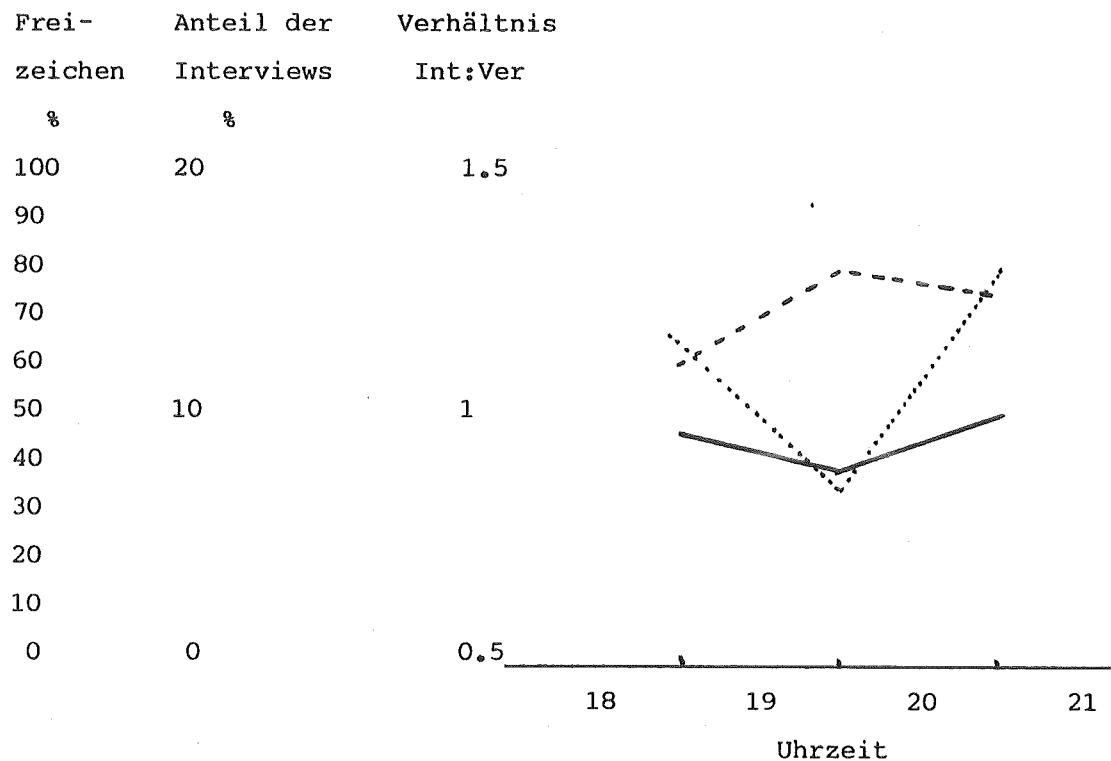


Abbildung 2:

Die Eignung der Abendstunden von 18 bis 21 Uhr zum Interview:

Freizeichen, Anteil der Interviews und Verhältnis Interviewer:Verweigerer



Legende: Freizeichen \_\_\_\_\_

Anteil der Interviews -----

Verhältnis Interviewer:Verweigerer .....

Der Anteil der Interviews zu verschiedenen Zeiten addiert sich nicht auf 100%, da weitere Interviews auch zu anderen Tageszeiten, vor allem am Wochendende, durchgeführt wurden.

Anmerkungen

1. Die Befragung wurde vom 14. - 26. März 1983 durchgeführt, ein Datum, das für die Interpretation der Ausschöpfung insofern von Bedeutung ist, als es der Höhepunkt der Diskussion um die für den Mai 1983 geplante Volkszählung war. Bei den hier berichteten Untersuchungen betrug die Ausschöpfung, d. h. der Anteil der realisierten Interviews von der Gesamtzahl aller bearbeitbaren Telefonnummern, 45% (während der Volkszählungsdebatte) bis 65%. Durch Intensivierung der Bemühungen, z.B. Rückruf nach ursprünglicher Verweigerung, ist diese Quote erhöhbar.
2. Abschnitt 4 wurde von S. Hormuth auf dem 34. Kongreß der Deutschen Gesellschaft für Psychologie vorgetragen, Abschnitt 2 und 3 ebenfalls von S. Hormuth vor der Methodensektion beim 22. Deutschen Soziologentag. Kurzfassungen finden sich im Bericht vom Kongreß der DGfPs (D. Albert, Herausgeber; Verlag Hogrefe) und im Materialienband vom Soziologentag (H. Franz, Herausgeber; Westdeutscher Verlag), der auch einen Beitrag von E. Brückner zur Kommunikationssituation enthält, aus dem einige Argumente im vorliegenden Beitrag enthalten sind.
3. Die Untersuchung wurde in den USA durchgeführt, während der erste Autor als Post-Doctoral Fellow an der Northwestern University vom National Institute of Mental Health gefördert wurde. Die deutsche Replikation der Untersuchung wurde von ZUMA durchgeführt (Brückner, Hormuth & Sagawe, 1982).

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: BRAUNMÜHL, 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 sprach-psychologische 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.: Möglichkeiten und Grenzen der theoretischen Aussagekraft von multidimensionalen Skalierungen bei der Untersuchung menschlicher Informationsverarbeitung. Teil I: Formale und wissenschaftstheoretische 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 Rekonstruktion 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 Theorien von Frauen über ihre erste Schwangerschaft: Theoretische Konzepte 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 Hypothesenkorpus. August 1980

Diskussionspapier Nr. 22: TREIBER, B.: Erklärung von Förderungseffekten in Schulklassen durch Merkmale subjektiver Unterrichtstheorien ihrer Lehrer. Oktober 1980

Diskussionspapier Nr. 23: RÖHRLE, B. & KOMMER, D.: Handlungstheoretische Betrachtungen zur primären Prävention psychischer Störungen. Februar 1981

Diskussionspapier Nr. 24: VOIGT, F.: Die Entwicklung des Zahlbegriffs. Teil I: Entwicklungslinien des Zahlbegriffs im Vorschulalter: Übersicht über theoretische Probleme und empirische Untersuchungen, mit einer Bibliographie zur Zahlbegriffsentwicklung. Teil II: Entwicklungslinien des Zahlbegriffs im Vorschulalter: Deskriptive Untersuchung des kindlichen 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 Biologisten 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

Diskussionspapier Nr. 37: SCHNEIDER, G.: Reflexivität als Grenzproblem einer kognitiven Psychologie. August 1983

Diskussionspapier Nr. 38: GEUTER, U.: "Gleichschaltung" von oben? Universitätspolitische Strategien und Verhaltensweisen in der Psychologie während des Nationalsozialismus. Oktober 1983, Historische Reihe Nr. 11

Diskussionspapier Nr. 39: KRUSE, L.: Drehbücher für Verhaltens-  
schauplätze oder: Scripts für Settings. Dezember 1983

Diskussionspapier Nr. 40: GRAUMANN, C.F.: The individualisation  
of the social and the desocialisation of the individual -  
Floyd H. Allport's Contribution to Social Psychology -. Mai  
1984, Historische Reihe Nr. 11

Diskussionspapier Nr. 41: GRAUMANN, C.F. & KRUSE, L.: Environ-  
mental Psychology in Germany. November 1984