Studies in economic
psychology: 82

Kjell Nowak

THE PSYCHOLOGICAL STUDY OF MASS COMMUNICATION EFFECTS:

On the validity of laboratory experiments and an attempt to improve ecological validity

Economic Research Institute
Stockholm School of Economics
1972

Kjell Nowak

The Psychological Study of Mass Communication Effects:

On the validity of laboratory experiments and an attempt to improve ecological validity

Akademisk avhandling

som med vederbörligt tillstånd för avläggande av ekonomie doktorsexamen vid Handelshögskolan i Stockholm framlägges till offentlig granskning
måndagen den 8 maj 1972 kl. 10
i sal 205 å högskolan, Sveavägen 65, Stockholm.

Stockholm 1972

I have had many inspiring discussions over the years about methodological and theoretical problems and who has spent much time and effort giving me constructive criticism of previous versions of the report. I also deeply appreciate the intellectual stimulation and practical advice given to me by Professor Karl-Erik Wärneryd in the course of the study.

A major part of the report was translated by Mrs. Charlyn Hultén, Mrs. Lill Graham, Mrs. Kerstin Niklasson, Mrs. Harriet Lundh, and Mrs. Marlene Ölander have taken care of typing and production. My thanks go also to them. Last, but certainly not least, I wish to express gratitude and appreciation to my wife and children for their understanding and patience during the preparation of this report.

Stockholm, April, 1972.

Kjell Nowak

CONTENTS

Acknowledgements i

Table of contents iii

I. Introduction 1

I:1. The background 2

I:2. The validity of experiments 4

I:2a. Internal, external and ecological validity 5

I:2b. Generalizability and representativeness 7

I:2c. Criticisms of psychological experiments 8

II. Sources of invalidity in laboratory research on communication effects 10

II:1. The experiment as social interaction 11

II:1a. The experimenter 12

II:1b. The subject 15

Modes of responding 16

Deception 17

The college volunteer subject 19

II:1c. Conclusion 20

II:2. The social interaction in experiments on communication and attitude change 22

II:2a. Differential effect of "experimenter-as-communicator" 22

II:2b. Compliance or cognitive restructuring? 24

II:2c. Suspicion 25

II:2d. Conclusion 26

II:3. Measurement and analysis as sources of invalidity 27

II:3a. Measurement effects 27

II:3b. Tests of significance 30

II:3c. Analysis of change 33

II:3d. Conclusion 34

II:4. The research setting 35

II:4a. The laboratory and the "real world" 35
Mundane realism 35
Isolated processes 37

II:4b. The communication situation in the laboratory 39
Isolated processes 39
The size of effects 41

II:4c. Conclusion 43

II:5. Independent variables 45

II:5a. The S-O-R-R paradigm 45

II:5b. Irrelevant stimulus properties 48

II:5c. Conclusion 49

II:6. Dependent variables 51

II:6a. What is attitude change? 51

II:6b. Measures of attitude change 52

II:6c. Attitude change and behavior 54

II:6d. Conclusion 57

II:7. Alternative experimental settings 59

II:7a. Changing the subject's role 59

II:7b. Disguised experiments 62

III. A field experiment on the effects of message form 66

III:1. Conceptualization of the research problem 67

III:1a. The approach to operationalization of independent variables 68

III:1b. Design of the experimental treatments 69

III:1c. Dependent variables 73

III:2. Hypotheses and procedure 75

III:2a. Hypotheses 75

III:2b. Design of the experiment 76

III:2c. Data collection 77

III:3. Check of the experimental treatments 80

III:3a. The method used 80

III:3b. The choice of rating scales 82

III:3c. Intended differences between the program versions 83

III:3d. Results of the Mail questionnaire study 84
The forming of indices 84
Analysis of the indices 87

III:3e. Program ratings in the Effects study 90

III:4. The Effects study: problems of internal validity 94

III:4a. Comparing naturally existing groups 95

III:4b. Self-selection 98

III:4c. The present experiment 99

III:4d. Analysis of group comparability 101
Relationship between membership in experimental groups and variables indicating initial knowledge 101
Determinants and extent of self-selection 102
Comparison of the non-exposed groups 104
Other control variables 105

III:4e. The comparability of the experimental groups - summary 107

III:5. Results of the Effects study: Exposure behavior variables 108

III:5a. The exposure time variable 108

III:5b. The exposure activity variable 111

III:6. Results of the Effects study: Knowledge and attitudes 113

III:6a. The knowledge variable 113

III:6b. The attitude variables 116

III:6c. Different comparisons 120
Length of exposure 120
Between-sample and within-sample comparisons 121

III:6d. Results in the Low-education group 125
Analysis of the knowledge variable 126
Analysis of the attitude variables 126

III:6e. Results in the High education group 128
Analysis of the knowledge variable 128
Analysis of the attitude variables 130

III:6f. Summary of the comparison of the versions 130

III:7. The absolute effect of the programs 131

III:7a. The relationship between exposure and the dependent variables 132

III:7b. Comparisons between the exposed and non-exposed groups 135

III:7c. Analysis of the knowledge variable 138

III:7d. Summary 139

III:8. Discussion 141

III:8a. The pattern of results 141

III:8b. Power 143

III:8c. Design 145

III:8d. Concluding remarks 146

References 150

Appendices: 165

1. The Baseline study
2. The Incidental technique study
3. The Mail questionnaire study
4. The Producer study
5. The Effects study
6. The Information-seeking study

An appendix in Swedish is available, containing certain materials not included in this report.

I. INTRODUCTION

I:1. THE BACKGROUND

During the last 20 - 30 years there has been a considerable interest in attitude change following exposure to communication stimuli. A vast number of laboratory experiments have been conducted, and there has been considerable theoretical development on issues like selective exposure to information, perceptual and judgmental responses to communication stimuli, dynamics of cognitive-affective structures, and vicarious learning. Typical of these experiments is the use of one-way, often non-personal communication situations in which subjects are required to read, watch, or listen to a more or less persuasive message under some alleged purpose. The theoretical background of many of these studies is not specifically related to the effects of mass communication, but hypotheses and results are quite often implied to be applicable in that field. Studies of this nature are almost always included in reviews of research on mass communication effects, and there is no doubt that much of the available theory on the persuasive power of mass communication is based on social psychological laboratory research.

In the context of mass communication theory,laboratory research on effects of communication has been questioned on several grounds. One such argument is related to the general criticism of mass communication research as being too preoccupied with the notion of <u>effects</u> which, it is argued, not only has limited value for the understanding of mass communication as an important social institution, but also inhibits the development of a more appropriate theoretical framework (UNESCO, 1970). Another point of criticism is that effects research (not only in the laboratory) is limited to short term effects, usually at the level of individuals or small groups. Furthermore, this micro level analysis is considered to be

concerned more with effectiveness than effects of mass communication. The claim is that attention has been focused on effects intended by mass media communicators, and that much research has been geared to the interests of the mass media (McQuail, 1969; UNESCO, 1970).

These points are certainly important with respect to the relevance of the so-called effects research for the development of mass communication theory. The main purpose of the present study, however, is to consider another source of criticism against laboratory research on communication effects, viz. its validity. In recent years there has been a growing concern among social psychologists about laboratory procedures and practices from ethical, methodological, and philosophical points of view. The empirical investigation to be reported in later sections (a field experiment) is a consequence partly of several articles published by prominent researchers raising doubts about the value of laboratory research, partly of the author's own experiences from laboratory studies. Before reporting the attempt to become more "naturalistic", a review will be made on some important problems relating to the validity of laboratory research on communication and attitude change. To some extent it is necessary to consider the characteristics of psychological laboratory experiments in general.

First, however, some comments are required to clarify how the concept of validity is used in this report, as applied to an experiment or an area of research.

I:2. VALIDITY OF EXPERIMENTS

Experimentation is closely related to the concept of causality. Though causality is much debated as an empirical and philosophical concept, it is certainly a central part of much theoretical (and everyday) thinking. Blalock (1964) has emphasized the unsurmountable gap between theory and empirical research in the sense that causal relationships can never be demonstrated, whereas they not only are theoretically useful but almost impossible to avoid as assumptions about the real world.

Whereas causality cannot be demonstrated, causal models can be evaluated, i.e. empirically tested, provided that they are formulated as closed systems. The value of the test depends on the extent to which the phenomena observed may be assumed to have the same character of a closed system. Since it is impossible to show that no other variables than those included in the model are in operation at the empirical level, "a causal relationship between two variables cannot be empirically evaluated unless we can make certain simplifying assumptions about other variables" (Blalock, 1964, p. 13).

The basic principle of experimentation is to arrange a piece of reality in such a way that it resembles as much as possible the closed system of a theoretical causal model or, in other words, to reduce the number of necessary assumptions in evaluating the model. This is the idea of experimental control (including randomization), and the controls undertaken reflect the experimenter's theory or guesses about which variables are related to the phenomenon under investigation. The primary purpose of experimental control is to rule out plausible rival hypotheses (Campbell, 1969), i.e. to ensure that the independent variables are not confounded with other variables. (Another purpose, of course, is to reduce random error).

Logically, it can never be shown that confounding has not occurred, but on theoretical and empirical grounds it may be considered more or less probable. As knowledge and theory develop in a certain field, new plausible rival hypotheses appear, and new measures of control become necessary.

Experimental procedures and the accompanying controls thus must depend on the nature of the subject matter of research.

Psychological experiments have often been criticized for adhering too much to a methodology developed in the natural sciences, neglecting the specific problems inherent in human research. Sometimes the laboratory experiment as such is considered to be without value in psychological research:

> "If we review briefly the basic premises of the experiment as carried out in physics and chemistry, we will see this more closely. The first premise is that the sample for investigation will resemble the whole ... The second assumption is that the sample will not be changed by removal from the setting in which it occurs in nature, and that it will not be changed by moving it to the laboratory from the setting ... A third premise ... is that the experimenter remains an observer; that he does not participate in and, by his participation, affect the outcome of the experiment. The fourth premise is that all the variables, save that which we wish to experiment, can be controlled.
>
> Without belabouring the matter, let me say categorically that not one of these premises is valid when one comes to experiment with human behavior" (Cameron, 1963, p. 269, cited by Smith, 1970, italics in original).

The points raised in this quotation may be used as a basis for discussing different concepts of validity.

I:2a. Internal, external, and ecological validity

The last two points in the quotation refer to what has been called the internal validity of experiments (Campbell, 1957; Campbell & Stanley, 1963).

This refers to the question whether the experimental manipulations (the independent variables) actually can be assumed not to be confounded with extraneous variables. With a given research setting and procedure and given operational definitions of variables, is the observed effect (or lack of effect) due to influence (or lack of influence) from the independent variable, or are there plausible alternative explanations to the results obtained? Such alternative hypotheses may refer to the presence of external factors, co-varying with the experimental variables (e.g. experimenter expectancy), but also to the possibility that experimental manipulations have resulted in variations in variable W rather than variable X. Manipulations intended to affect, say, dissonance, might affect some other variable (e.g. evaluation apprehension). To the extent that such rival hypotheses cannot plausibly account for the outcome, the experiment is said to have internal validity.

The first two points, on the other hand, refer to the degree of correspondence between what is found in the experimental situation and what actually exists in the natural habitat. Using Campbell's terminology, it is a matter of the external validity of the experimental findings, i.e. the conditional character of the empirical relationship found in the laboratory and its existence under conditions which in various respects are different from those of the experiment. According to the quotation, results of psychological experiments cannot be generalized to naturally occurring situations since conditions are inevitably different.

Naturally, some degree of external validity is assumed in all research, but the meaning of "external" may vary according to the purposes of the research. Speaking of a whole area of investigation, external validity seems to come close to ecological validity (Brunswik, 1956) in a general sense, i.e. the generalizability of theories and findings to the natural habitat of the phenomena under study. For particular experiments, however, external validity may have a more restricted meaning, e.g. generalizability to other research situations, other subjects or experimenters, different operational definitions of variables, etc. Usually, a particular study can only aim at applicability in a very restricted set of naturally occurring situations, but if it is to fulfil any scientific purpose a basic requirement must be generalizability at least to other research settings.

I:2b. Generalizability and representativeness

Assuming internal validity of a piece of research that has demonstrated a certain relationship, a failure to find this relationship in another situation may have three kinds of reasons. One is simply lack of internal validity the second time. Another is that the relationship is contingent upon a certain condition, not specified in the theory, which was not present the second time or, in other words, that the independent variable interacts with an unspecified factor that was differentially operative on the two occasions. A third reason may be that the theoretically specified conditions do not exist or cannot be made to exist in the second situation. A relationship found in the laboratory may not be possible to demonstrate for such reasons in a less controlled setting.

In the second case the lack of generalizability is due to the fact that the theory or hypothesis is incompletely specified - one or more variables have to be included to allow generalization. In the third case, the postulated relationship exists and may be well specified as to interacting variables, but it cannot be demonstrated except under highly controlled conditions. A relationship which is well specified and is consistently demonstrated under highly controlled but varying conditions surely must be considered to have external validity, even if it cannot be demonstrated under less controlled conditions. It certainly has a value for developing a theory, and since this is a primary goal for basic research, external validity in this sense seems to be all that is called for by purely scientific standards. In other words, the fact that a theoretical relationship cannot be demonstrated in a natural situation is in itself not a sign of low external validity.

The claim for ecological validity is a step further and requires that relationships, variables, and processes studied be ecologically representative of the field to which a theory applies.[1] Any theory states or implies that under such-and-such conditions a certain outcome is expected, and to the extent that these conditions are representative of the natural habitat, the theory can be said to have ecological validity.

---

1) It should be noted that ecological validity is not used here in the rather specific meaning given to the term in the context of the lens model (cf. Hammond, 1966; Björkman, 1968).

The meaning of "natural habitat" is not very clear, of course, in many fields of the social sciences, and it may very well be subject to changes, spontaneously or through social action. Ecological validity can therefore not be a once-and-for-all given characteristic of a theory.

Validity of experiments as well as validity of measurement must be judged relative to a purpose (Kaplan, 1964). The meaning of internal validity is rather clearcut in this respect. The purpose implied is that of demonstrating the existence (or non-existence, or form, or strength) of a (causal) relationship between certain variables, under more or less specified conditions. The question is whether the empirical results may be interpreted as such a demonstration.

External validity, on the other hand, is a more ambiguous term. It has to do with generalizability of results, but for what purposes? Since every researcher may have his own goals as to the generalizability of results, it seems appropriate to use the term external validity with reference to the purpose of a piece of research as defined by the investigator. Ecological validity, on the other hand, is better used with reference to the purpose of a theory or a whole body of research. At this level, representativeness of phenomena and relationships is assumed to be an important objective which is related to general notions of the role and function of basic research. (This is further developed in section II:4.) Defined in this way, ecological validity does not follow from external validity - a certain relationship may be highly generalizable under specified conditions but still be non-representative or have limited significance in naturally occurring situations. The reverse is also true - relationships in a theory may be ecologically valid but still not be generalizable, e.g. because they are insufficiently specified as regards their conditional character.

I:2c. Criticisms of psychological experiments

Laboratory experimentation in psychology has been more or less severely

criticized on the basis of each of these three criteria for validation. The most serious question, of course, is that of internal validity. During the 1960's a good deal of research has been published which rather clearly indicates that extended controls and changes in procedures may be necessary in at least certain types of experiments. Common to all these studies is an emphasis on the social nature of the experimental situation, and the importance of the experimenter-subject interaction. Although this aspect of human research is a long recognized fact (see Rosenzweig, 1933), it has not received much attention in texts on experimental methodology and design.

The concern with external and ecological validity is reflected in two kinds of argument. One is the claim that experiments should be much more frequently replicated, both to rule out the possibility of chance results and to establish generality across a variety of research situations. Implied in this argument seems to be that the ecological limitations of laboratory experiments as such are accepted, but that there is doubt about external validity so that the results need more corroboration. The other kind of argument focuses on the artificial character of the laboratory, and stresses the need for more situational realism, less reactive methods of observation and measurement, and more representative samples of subjects. Here the emphasis seems to be more towards representativeness and generalizability to naturally occurring situations.

The following section (II:1) briefly reviews certain characteristics of psychological laboratory experiments which have been demonstrated to operate as threats to internal validity or generality across research settings (external validity). These possible artefacts are then discussed more specifically in the context of experiments on communication effects (II:2). Later, the focus will be more on ecological validity of research on communication and attitude change.

II. SOURCES OF INVALIDITY IN LABORATORY RESEARCH ON COMMUNICATION EFFECTS

II:1. THE EXPERIMENT AS SOCIAL INTERACTION

As several authors have pointed out, the procedure and conceptualization of the typical laboratory experiment in psychology is based on ideas derived from 19th century philosophy and research in the natural sciences (Luchins & Luchins, 1965). The ideal experimental setting is a situation where the experimenter acts as a neutral observer, not affecting and not affected by the object under study, and where the subject, initially neutral to the situation as such, is a naively (naturally) responding organism, reacting primarily to those characteristics of the situation which the experimenter controls (the stimuli). That this is an unattainable ideal is evident, but the widespread use of laboratory experiments reflects a confidence that available techniques of measurement and situational control can ensure a solely random or assessable influence of extraneous factors on the process under study. In recent years, however, growing doubts have been expressed among psychologists about the adequacy of traditional laboratory procedures and their underlying assumptions. A vast number of studies have demonstrated various effects of the social nature of laboratory experiments, showing both that the experimenter cannot be considered a neutral observer and that he unintentionally and subtly may systematically bias subjects' responses, and that subjects are by no means passively responding to well-defined stimuli but that rather they are actively trying to interpret the situation as a whole, including the experimenter's intent and expectations.

One line of research on the characteristics of the laboratory experiment, most prominently represented by Robert Rosenthal and his group, has concentrated on experimenter expectancy effects on the dependent variables. To some extent this approach is a parallel to earlier studies on

interviewer bias in field surveys (e.g. Hyman, 1954), and on examiner differences in clinical or testing situations (see Friedman, 1967; Kintz et al, 1964). Another, more heterogeneous area of research is concerned with the validity of the prevailing view of the subject "as object" and the experimenter-subject relationship as that of "person-to-thing, with its attendant tendencies of domination, manipulation, and control" (Schultz, 1969, p. 217). The criticism of laboratory experiments emanating from these studies is based partly on considerations of internal or external validity, partly on ethical or philosophical grounds. Some critics seem to see the solution in new and extended measures of control (e.g. Orne, 1969; McGuire, 1969; Rosenberg, 1969), whereas others propose more radical changes in experimental tasks and the experimenter-subject relationship (e.g. Kelman, 1968; Argyris, 1968; Jourard, 1968).

II:1a. The experimenter

It is well known that experimenter (or interviewer or examiner) differences may be an important source of variation in research results. Sex, age, and race of the experimenter often influence the outcome, but it is not clear to what extent this reflects different behavior among experimenters and to what extent subjects react differentially depending on experimenter characteristics. Results cited by Rosenthal (1969) indicate that both phenomena may be of importance.

Characteristics like sex and race of experimenter are usually paid attention to in the sense that they are reported in the published studies, but they are more seldom discussed in relation to results obtained. Personality characteristics, on the other hand, are very seldom reported, although they have been shown to be related to subjects responses (Rosenthal, 1966). From the internal validity point of view, these main effects of experimenter characteristics are most important in cases where different experimenters are used in the various experimental con-

ditions or where experimenter characteristics may be assumed to interact with the independent variables. The first case is not very frequent, but the other is a threat to internal validity which is specific to each single study and which stresses the need for independent replication.

Of much more general concern are the effects of experimenter expectancy, which have been impressively demonstrated by Rosenthal's research (summarized in Rosenthal, 1966 and 1969). Rosenthal's hypothesis, which has received considerable support across a variety of experimental tasks, subjects (including animals), researchers, and experimental settings, is that the experimenter may unintentionally and subtly influence subjects' responses in a direction confirming his expectations, i.e. in line with his research hypothesis. In an analysis of a number of studies on such experimenter bias, Barber & Silver (1968a and b) have argued that there was little evidence that the effects were really due to unintentional cues from the experimenter, and that the results may have been a consequence of observation errors or intentional cheating on the part of the experimenters studied (who were in many cases undergraduate students). Still, it seems undeniable that such factors cannot explain away the total evidence of experimenter expectancy effects, nor can a mechanism of operant conditioning (Rosenthal, 1968; 1969).

On the other hand, in spite of careful analysis of films and recordings of experimenter-subject interactions, "no well-specified system of unintentional cueing has been uncovered" (Rosenthal, 1969, p. 252). Both auditory and visual cues seem to be of importance, but it is not clear in what ways. It has been suggested that the experimenter in the course of the experiment unconsciously learns through reinforcement which cues affect subjects' behavior in a certain direction, but there is no unequivocal support for this hypothesis.

The demonstration of experimenter expectancy effects no doubt points to an important problem in psychological research, but one must bear in mind that most of the studies are designed to invite expectancy effects (Aronson & Carlsmith, 1968). The experimenters in these studies

usually run only one experimental condition, which is of course seldom the case in actual laboratory research since it would mean a confounding of experimenter characteristics with the independent variable. Although bias has been observed also when each experimenter runs all conditions, the effects are much less dramatic (Rosenthal, 1969). In studies where all conditions are run simultaneously (group experiments), which is often possible in research on communication effects using written materials, the experimenter expectancy effect is probably less serious, since the experimenter cannot as easily give different cues to subjects in different conditions.

There seems to be no universal way to avoid experimenter expectancy bias. The principal investigator's withholding information from the experimenter as to the purpose of the study would extend the problem of "demand characteristics" (see below) to include the experimenter as well as the subjects. Besides, there is evidence that the principal investigator may unintentionally and covertly communicate his own expectations to his assistants even when he did not directly inform them about the research hypothesis (Rosenthal et al, 1963). Tape recorded instructions in the absence of the experimenter is another solution suggested, but this may create more problems than it solves, since the situation will be more ambiguous from the subjects' point of view. In principle, the experimenter should never know which experimental condition he is running, but this is usually practically impossible. Aronson & Carlsmith (1968) have suggested that the experimenter should be ignorant up to the point where instructions, materials or the situation differs between conditions. At that point he could by some simple device randomly decide the conditions to be run. It may be added that if, after that, the rest of the study is "automated", e.g. by means of tape recorders or films, experimenter expectancy effects would be rather improbable.

Even if the problem of experimenter expectancy effects is sometimes overstated, it is clear that the "standardization myth" (Friedman, 1967) of textbooks on experimental design and procedures is ripe to be replaced by more concrete warnings, norms, and advice as to the experimenter's behavior and role as a significant stimulus object and

possible confounding variable (McGuigan, 1963). The experimenter's unintentional influence on subjects' responses seems to deserve as much attention in the planning of experiments as conventionally accepted sources of unintended systematic variation in the dependent variables, especially in social psychological research where procedures are often very complex (cf. Chapanis & Chapanis, 1964).

II:1b. The subject

The social relationship between subject and experimenter has an asymmetrical character which, according to Argyris (1968), is comparable to a management-employee situation. The subject has implicitly agreed to come under the control of the experimenter and to do what he is told without asking why. Thus there is an unequal distribution of power and information (Riecken, 1962). The passivity and ignorance of the subject, as well as the common instruction to respond spontaneously, are integral parts of the prevailing ideas about experimental control. They constitute attempts to make the experimental situation a closed system (Sjöberg, 1971a) and to approach a natural, naive mode of responding on the part of the subject.

It is evident, though, that subjects cannot be assumed to be neutral to an experiment - they initially have certain expectations, beliefs and attitudes to psychological experiments in general and to the study in question, and they may utilize any cues available, from the moment of first knowing about the study, to form hypotheses about its purpose and to define their own role in relation to the experimental situation. Subjects respond not only to the stimuli as defined by the investigator, but to a whole context which has certain "demand characteristics" (Orne, 1962). These demands refer to any cues in the experimental setting which affect the subject's perception of the situation, but even the mere awareness of being in an experiment may significantly affect responses on the dependent variables (Silverman, 1968).

Modes of responding. Various suggestions have been made as to the typical mode of responding among college subjects to the demand characteristics of the experimental setting. Orne (1962) proposes that subjects are mainly concerned that their behavior serves to validate the research hypothesis, as perceived by them. He feels that college subjects willingly accept the social role assigned to them as subjects, having full confidence that the experimenter will do them no harm and that the operations required by him are important and meaningful. Orne's (1962; 1969) demonstrations of the docility and cooperativeness of subjects are very convincing in this respect, as are Milgram's (1963; 1965) studies of obedience, utilizing normal adult samples.

Fillenbaum (1966) and Holmes (1967) give evidence that the more experienced subjects, assumedly more aware of deception practices, do not play the role of the over-cooperative "good subject" as found in Orne's studies, but rather the role of the "faithful subject", adhering as much as possible to the instructions and not trying to figure out the purpose of the research. According to Argyris (1968), college subjects nowadays come to an experiment fully prepared to be deceived, resenting compulsory participation and reacting with counter-strategies characterized by withdrawal or active attempts to distort the outcome of the study. Similar ideas have been expressed by Masling (1966) who talks about the "screw you effect".

Riecken (1962) emphasizes the tendency of a subject to "put his best foot forward", i.e. to present himself in as good a light as possible. A similar line of thought has been developed by Rosenberg (1965; 1969). He has empirically demonstrated that cues leading the subject to expect that he is being evaluated may affect results to a significant degree, and that such cues may be confounded with the experimental manipulations of certain types of studies. This mode of responding, called "evaluation apprehension" (Rosenberg, 1965; 1969), may also be related to experimenter expectancy effects - one study (Minor, 1970) showed that such effects could be observed only in a condition where subjects were personally concerned with their performance. It has also been demonstrated that evaluation apprehension may be a more salient motive than complying with demands (Sigall et al, 1970).

Deception. The whole issue of demand characteristics and the subject's attempts to structure and adapt to the situation is closely related to the common practice of deceiving subjects as to the true purpose of the study. It is somewhat paradoxical that deception, which is used to reduce demand cues and assure equal expectations or modes of responding across conditions, in itself may counter-act these objectives. Deception or not, the subjects will always respond to the stimuli and the setting as perceived by them, but, as pointed out by Argyris (1968), the question is under what conditions the investigator has the greatest awareness of and control over what subjects actually respond to. It has certainly not been shown that deception is superior in this respect.

However, deception practices are very widespread in certain areas of research - one survey indicated that about 20 per cent of all studies and 80 per cent of conformity studies published in four major journals used deception (Stricker, 1967) - and as long as they are used there will always be the question whether they were successful or not and whether they were equally successful in all conditions. In one area of research this has been a focus of controversy for many years, viz. in verbal conditioning, where the problem of awareness of reinforcement contingencies is a central issue of validity (see, e.g. Greenspoon & Brownstein, 1967; Spielberger, 1965). In other areas, it is only in recent years that concern for the effectiveness of deception and the effects of subject suspiciousness (or awareness) has become more widespread. Surprisingly little attention has been given to these questions in published results (Stricker, 1967) in view of the importance of the assumption of naive responses for generality of results.

A review of published studies (Stricker, 1967) indicates that subject suspiciousness is not very often assessed and that, when this is the case, very few subjects are classified as having been suspicious about the purpose of the study or the procedure of deception. Orne (1962) has suggested that the ordinary post-experimental interview may be conducted in a "pact of ignorance", where the demand characteristics are such that subjects are led not to reveal their suspicions. Data consistent with this hypothesis have been presented by Levy (1967) and Goldberg & Lichtenstein (1970). The reasons given to the subject

for conducting the post-experimental interview seemed to affect somewhat the rate of confession (see also Rubin & Moore, 1971), but very few subjects (10 - 15 per cent) among those who were actually fully informed (through a "tipping off" procedure before the experiment), revealed this in the interview.

Suspiciousness seems to be related to personality variables (Stricker et al, 1967; Rubin & Moore, 1971), training in psychology (Rubin & Moore, 1971), and previous experimental history (Cook et al, 1970; Holmes & Applebaum, 1970; Silverman et al, 1970). A number of studies have tried to clarify the relationship between previous experience of deception and performance in later experiments, but the results are not very consistent. Tentative conclusions would be that performance is not affected unless there is a high degree of similarity between the experiments (Brock & Becker, 1966; Fillenbaum, 1966; Stricker et al, 1967; Stricker et al, 1969), or unless the research hypothesis can be rather easily guessed in the subsequent experiment (Cook et al, 1970). The latter study also indicates the possibility of a curvilinear relationship between degree of suspiciousness and effect on performance - bias is least likely when demand characteristics are either very obvious or very unobvious in view of the subject's previous experience (Cook et al, 1970; see also Orne, 1962). Similar findings were reported by Golding & Lichtenstein (1970), who found that naive and fully informed subjects produced comparable results. As they point out, however, the meaning of the responses is totally different in the two groups. A curvilinear relationship is consistent with the reasonable assumption that biased results are most probable among subjects who are suspicious but feel very uncertain about the true purpose and what is expected of them.

Where suspicion or previous deception has been related to performance, the effect is generally towards less conformity or less yielding to demand characteristics (Rubin & Moore, 1971; Stricker et al, 1967; Allen, 1966a; Cook et al, 1970; Holmes, 1967), but also towards more favorable self-presentation (evaluation apprehension; Silverman et al, 1970). Suspicion may further interact with personality traits in its effect on performance - Rubin & Moore (1971) found that suspicious

subjects scoring high on the F scale tended to yield more to demand characteristics, whereas suspicious subjects scoring low on the F scale showed less conformity than non-suspicious controls.

Results obtained so far are thus rather complex, and there is no evidence of a dominant mode of responding as a consequence of prior deception or suspiciousness. Neither the "good" or the "faithful" subject, nor the "negativistic" subject seems to be typical. It is still an open question under what conditions the practice of deception is a threat to internal or external validity. Clearly, suspicion can affect performance, but there is no basis for conclusions as to how common this is in ordinary theory-testing experiments (Cook et al, 1970).

The college volunteer subject. Closely related to the problems of demand characteristics and effects of deception or suspiciousness is the heavy reliance on college samples in psychological laboratory research. A review of published research in four major American journals (Schultz, 1969) showed that 70 - 80 per cent of the studies used college subjects, and about half of them introductory psychology students.

Even if these samples had been randomly drawn they would give rise to concern about generalizability: although it is often stated that the population sampled is of no importance as long as its characteristics do not interact with the independent variables, the probability of such interactions is practically unknown, since very few studies are carried out with normal, adult samples. Present practices seem to indicate considerable disinterest in the fact that inferences about relationships or differences are contingent upon the characteristics of the population sampled.

However, many results are not only limited in their generality to psychology students, but to students who volunteer to participate and who also keep their appointments. Two comprehensive reviews (Rosenthal & Rosnow, 1969; Rosnow & Rosenthal, 1970) show that volunteer subjects do differ from non-volunteers on a number of social and psychological variables, e.g. need for approval, intelligence, and authoritarianism. In certain areas of research there is reason to suspect that such

variables interact with experimental treatments and there are a few studies which have attempted to investigate whether volunteers actually respond differently from non-volunteers (Horowitz, 1969; Rosnow & Rosenthal, 1966; Rosnow et al, 1969; Hood & Back, 1971; Rosnow & Suls, 1970). Judging from these studies, a tentative hypothesis would be that volunteers are more sensitive to or more complying with demand characteristics. Rosnow & Rosenthal (1970) propose, though, that the effect of volunteer status may be pronounced only when demand characteristics are neither very obvious nor quite non-obvious. As was mentioned above, this point of uncertainty about demands is assumed to be crucial also as regards the effects of previous experiences of deception or suspiciousness of intent, but the more typical result among suspicious subjects was less compliance. In this connection it is interesting to note that, in an opinion change study, a pretest (which may be considered a medium obvious demand cue) turned out to dampen the persuasive impact among non-volunteers, whereas it had the opposite effect among volunteers (Rosnow & Suls, 1970). Volunteer status may thus affect dependent variables both directly (volunteers may respond differently to experimental treatments in certain kinds of studies) and indirectly (through responses to demand characteristics).

II:1c. Conclusion

The preceding review of research indicates that internal as well as external validity may be jeopardized by the social interaction in the laboratory situation. As regards internal validity, it seems to be a real possibility that experimental results may be more a function of experimenter expectancy, compliance, or evaluation apprehension than of whatever psychological process the study is assumed to investigate. Such sources of invalidity, more or less inherent in the laboratory procedure, may also be difficult to discover in the course of replication.

It is not known, however, how probable or how strong these artifactual effects are in various areas of research - the studies reviewed have been largely limited to simple demonstrations that such effects are possible. One exception is Rosenberg's (1969) research on evaluation apprehension as a confounded variable in certain experiments on cognitive dissonance. On the whole, though, specific results or generalizations have not so far been shown to be invalid due to such artifacts as discussed above. In fact, the studies demonstrating the possibility of artifacts may themselves be subject to exactly the same kind of criticism. However, although the considerable body of systematic knowledge available from experimental data cannot be rejected as invalid on the basis of the research discussed, the social nature of the experiment certainly deserves much attention and calls for considerable ingenuity, fantasy, and creative thinking among investigators. It is quite conceivable that psychological theories will be developed that take into account and include the characteristics of the interaction between experimenter and subject. In that case, the threats to validity discussed above might form essential parts of the theory.

As to external validity, on the other hand, the research reviewed gives rise to concern. This may be a less serious problem, since the fact that results are valid only under certain conditions is more likely to be discovered in the course of replication and development of theory. It seems evident that what has been called constructive replication (Lykken, 1968) or transition experiments (Campbell, 1957) should be given a much more prominent place in research and publication decisions.

II:2. THE SOCIAL INTERACTION IN EXPERIMENTS ON COMMUNICATION AND ATTITUDE CHANGE

The social interaction aspect of the laboratory situation is very probably of particular importance in studies of communication and attitude change. The experimenter is a kind of pseudo-communicator with respect to the experimental communication, and the subject must be assumed to relate its content and source to the experimenter, his instructions and perceived purposes. Accordingly, the impact of the experimental communication will be subject to two kinds of source effects - one stemming from the experimenter, the other from the alleged source of the message. In a sense, any laboratory experiment on communication and attitude change may be viewed also as a demonstration of "experimenter-as-communicator" effects.

II:2a. Differential effect of "experimenter-as-communicator"

Since the experimenter, as viewed by the (college) subject, presumably has the characteristics of a "positive" communicator (power, status, expertness, trustworthiness, etc.) and since the subjects are in a situation where source effects are expected (uncertainty, dependence, need for approval, etc.), the persuasive impact of an experimental message is most probably especially strong in the laboratory. This has often been pointed out (e.g. Hovland, 1959; Sherif et al, 1966), but usually with reference to the forced exposure characteristic of

laboratory studies (see below). Attempts to magnify effects is a common goal of laboratory procedures, but it seems probable that the experimenter effect may also be differentially strong or take different directions in the various experimental conditions.

One such possibility can be derived from certain hypotheses in theories of cognitive consistency (see Nowak et al, 1966, ch. 9; Feather, 1967). The occurrence and direction of a source effect is assumed to depend on whether the receiver perceives the source as being responsible for or in agreement with the standpoint advocated in the message. In most cases the "real" source is probably perceived in this way, but not necessarily a pseudo-communicator. It can be hypothesized that if the experimenter is positively evaluated and is believed to agree with the experimental communication, its impact would increase, but if he is believed to disagree the effect would be in the opposite direction. Some data from a study of volunteer bias give some support to this possibility of differential experimenter influence on communication impact (Rosnow & Rosenthal, 1966). As was mentioned above, volunteer subjects generally tend to be more sensitive than non-volunteers to demand characteristics. In this study, however, such an effect of volunteer status occurred only in the condition where subjects had reason to believe that the experimenter endorsed the experimental communication (which was anti fraternities). In the pro fraternities condition, volunteers showed less attitude change than non-volunteers. Assuming that volunteers are more sensitive to demand characteristics, among which the experimenter's "source properties" are essential, this interaction is in perfect accord with consistency theory. It would seem that interpretation of studies on attitude change as a result of communication requires knowledge of the subjects' attitude towards the experimenter and their beliefs as to his standpoint relative to the experimental messages.

Similarly, it is conceivable that in experiments on source effects (e.g. Hovland et al, 1953), the experimenter is perceived to be more in agreement with "credible" or "positive" sources than with the sources in low credibility conditions. For example, it is reasonable that students believe the experimenter to be more in agreement with

an experienced, sincere judge than with a delinquent charged with drug-peddling (Kelman & Hovland, 1953). The implication is that the experimenter may take on the role of temporary reference person, whose attitudes as perceived by the subjects will be related by the latter to the experimental message and its source.

II:2b. Compliance or cognitive restructuring?

Using the terms suggested by Kelman (1958) the kind of experimenter-as-communicator effect mentioned above would represent the process of influence called identification. Considering the nature of the experimenter-subject relationship, the experimenter effect may more often take place through compliance, which is expected to occur as a function of the power of the communicator and his ability to observe the receiver's behavior. It should therefore be most pronounced under non-anonymous conditions. Silverman (1968) found that the mere awareness among subjects that they were taking part in an experiment increased the persuasive impact of a short message, but only for those subjects who were non-anonymous (cf. Kelman, 1958). In reports on studies on communication and attitude change it is seldom stated whether data were obtained anonymously or not, but the prevalence of before-after designs, where names are usually necessary, indicates that non-anonymous responses are common.

In many cases compliance to demand characteristics is only a threat to the external validity in the sense that treatment effects are exaggerated, but in research on attitude change this is clearly not the case. Those theories most often applied to the problem of communication and attitude change (consistency theories and the social judgment theory) are not concerned with a process of overt compliance to persuasive pressures, but with perceptual and judgmental processes. To the extent that results of experiments are a consequence of compliance to the experimenter as a powerful person, they will have

little to do with the psychological processes which the investigator thinks he is studying. In other words, even if compliance to demand characteristics is equally strong in all experimental conditions, it is a threat to internal validity.

Compliance may also be differentially operative depending on personality characteristics of the subjects. In attitude change research variables like anxiety, dogmatism, self-esteem and sex have been found to be important sources of variation, and they form part of several hypotheses about persuasion (cf. Nowak et al, 1966). It makes a good deal of difference for theory development if these variables are mainly related to overt compliance in a situation of dependence or if they are really related to the dynamics of cognitive-affective structures.

II:2c. Suspicion

As was mentioned above, a general tendency in studies of the effects of suspicion has been a decrease in conformity and yielding to demand characteristics. In attitude change research there have been some studies especially designed to test the effect of audience awareness that the communicator is trying to exert influence in a certain direction (see Nowak et al, 1966; McGuire, 1966). Results indicate that awareness of persuasive intent may both increase and decrease the persuasive impact, depending on communicator characteristics and the receiver's initial position. McGuire (1969) has reviewed these and a whole range of other studies on communication and attitude change, trying to assess the possible effects of suspicion of experimenter's intent. He concludes that "the results of these many lines of research considered ... are not particularly alarming as regards the possible artefactual nature of results obtained under conditions that might make the subject suspicious of the experimenter's intent" (McGuire, 1969, p. 35). In most cases suspicion

had no effect; in some cases there was either an increase or a decrease in persuasive impact. The direction of the effect may depend on what the subject is suspicious about - if he feels that the experimenter expects a change he may yield to the demand, but if he feels that the purpose of the study is to test persuasibility he may show resistance to change (evaluation apprehension).

II:2d. Conclusion

In conclusion, the social interaction of the laboratory situation seems to be inevitably confounded with the process under study in experiments on communication and attitude change, i.e. with the symbolic interaction between the subjects and the alleged source of the experimental message. To what extent this is a threat to internal or external validity is not known but the problem certainly deserves attention, especially as regards the artifactual nature of compliance which may mean that "a distorted view of attitude change often emerges from the laboratory" (Weick, 1967, p. 53).

It seems very probable that the "experimenter-as-communicator" effect generally results in an overall exaggeration of persuasive impact, except when subjects feel aggressive towards the experimenter or believe that their persuasibility is being tested. The possibility that this effect operates differentially in the various experimental conditions may to some extent depend on the kind of design used in the experiment. When comparisons are being made between a control group (exposed to irrelevant information) and an experimental group, the experimenter effect on the dependent variable is essentially limited to the latter. On the other hand, where treatments consist of messages which represent different levels or values of one or more independent variables (say, complexity, familiarity, discrepancy) the experimenter effect is more probably equally strong in all conditions. Even in those cases, of course, there is the risk of interaction between characteristics of the message and the perceived attitudes of the experimenter.

II:3. MEASUREMENT AND ANALYSIS AS SOURCES OF INVALIDITY

The points raised so far refer to problems of internal and external validity which are closely related to the conventional procedures in social psychological experimentation in general. Another such general issue is the role of measurement procedures as a source of artifactual results, and the place of statistical analysis in the discussion of validity.

II:3a. Measurement effects

The influence of the act of observation on the experiment's outcome is one of the most evident threats to validity, and not only in psychology. Measurement procedures are probably an important part of the demand characteristics, and the subject's mode of responding is no doubt affected by the nature of the measurement used.

One issue that has been discussed rather at length is the effect of pretesting in before-after designs (Solomon, 1949; Campbell, 1957; Lana, 1964). A main effect of a pretest is a minor problem in theory-testing experiments, since the means of both control and experimental groups are equally affected, but the problem is the possibility of an interaction between pretest and treatments. Several studies on communication and attitude change, however, have consistently failed to demonstrate such pretest sensitization. In some

cases (where the communication was two-sided) a main effect has been found. "The overwhelming lack of a pretest sensitization effect when the pretest is used to measure existing opinions or attitudes is as convincing a demonstration as one is likely to find in social psychological research" (Lana, 1969, p. 134).

A much more difficult problem is posttest sensitization, i.e. the reactive nature of most dependent variable measures in the laboratory. It is evident that the measurement procedure may act as a demand characteristic, indicating to the subject what the experiment was about and what the experimenter expects, but apart from this, the measurement may actually affect the psychological process under study (Aronson & Carlsmith, 1968). In a communication experiment, the message presented could leave the subject completely indifferent, but upon seeing the opinion statements he very likely starts reflecting on the communication, remembering things he did not think of before, etc. The difficulty with posttest sensitization is of course that its effects cannot be assessed, and it seems doubtful whether results on pretest sensitization can be generalized to posttests also.

Both pre- and posttest sensitization refer to the measurement of the dependent variables. In many experiments on communication effects there is also a measurement of the independent variable, with the purpose of checking whether manipulations have been successful and to make possible internal analyses of the data. E.g. studies of fear arousal and attitude change usually include certain questions or ratings as to how the message was perceived, whether the subject felt fearful, etc. At least one such study (Wuebben, 1968) has shown a considerable influence of this procedure on the outcome. The dependent variable in that case was overt behavior (calling a doctor for an appointment), and it turned out that without a measure of perceived fear the low threat condition produced more calls, whereas with such a measure included it was less effective than the high threat condition. Since the practice of measuring independent variables is rather common, there seems to be need for further investigations of its possible consequences as regards validity.

An entirely different source of invalidity relating to measurement has been pointed out by Campbell (1969). Just as every experimental treatment contains any number of theoretically irrelevant aspects, which may be the actual explanation of its effect, "every measurement device is dimensionally complex with theoretically irrelevant vehicular components. The measured effects of the treatment could be due to one of these irrelevancies" (Campbell, 1969, p.362). In other words, the treatments might affect only an aspect of e.g. an attitude scale which is theoretically irrelevant. In some cases, irrelevant aspects of the measure may be controlled for (e.g. social desirability), but in laboratory experiments this problem has not received much attention.

Finally, there is the question of validity of measurement as such. In the literature on experimental artifacts this is a neglected problem, but at least as regards studies on communication effects measuring instruments are often of a clearly ad hoc character. This aspect of validity will be given some consideration in a subsequent section.

The most obvious solution to the artifactual effects of measurement sensitization is of course use of non-reactive measures (Webb et al, 1966). In certain areas of research that is no doubt possible, even in the laboratory, but in other situations theoretical concepts may be difficult to operationalize except in terms of structured verbal responses to specified stimuli (e.g. structural theories of cognition). Also, non-reactive measures tend to be crude, not allowing more refined concepts or analyses. There are, however, many cases where non-reactive measures in the laboratory have been successfully used (cf. Aronson & Carlsmith, 1968; Campbell, 1969), and where this is not feasible there is clearly a need for more extensive use of multiple dependent variables, which is one way of reducing or detecting invalidity due to measurement procedures (Campbell & Fiske, 1959).

II:3b. Tests of significance

During the last 15 years the significance test issue has been under debate in psychological as well as sociological journals (see Morrison & Henkel, 1970). One part of the criticism has been concerned with the underlying assumptions of the tests and the extent to which these are or may be neglected by investigators (Camilleri, 1962; Morrison & Henkel, 1969; Kish, 1959; Selvin, 1957; Gold, 1969; Winch & Campbell, 1969). Mainly, the emphasis has been on field studies and on problems like significance testing on non-random samples or whole populations, picking significant results from a large number of tests on the same data, and applying formulas assuming simple random sampling when the data have not been so generated. Certain practices also in the analysis of laboratory data have been attacked on the ground that they violate the principle of randomization, e.g. discarding subjects, redefining the independent variable according to data in the experiment (internal analyses), and testing of post hoc hypotheses (Chapanis & Chapanis, 1964; Aronson & Carlsmith, 1968; Meehl, 1967). Meehl (1967) has also pointed to the methodological paradox that in most psychological research improved power in statistical design leads to a prior probability approaching $\frac{1}{2}$ of finding a significant difference.

It is evident that problems such as those indicated above are relevant to the question of validity, but a more generally applicable basis for criticism of tests of significance is the possible confusion, in basic research and theory development, of statistical and substantive significance, or between statistical and scientific inference. It seems to be generally agreed that tests of significance are of limited importance in scientific inference, and it is even argued that they may be misleading and inhibit theory development (Morrison & Henkel, 1970).

The null hypothesis significance test in the typical laboratory experiment is simply one aspect of internal validity. The test permits the investigator to assess the probability that an observed relationship or difference is due to chance rather than the experi-

mental treatment. If this probability is considered to be sufficiently low, the next question must be "what does it mean?", and at this point the probability obtained in the significance test is irrelevant. The process of ruling out plausible alternative hypotheses, assessing under what conditions the result was observed, and judging its generality and theoretical relevance, all of which is essential to scientific inference, lies outside the realm of statistical significance. Nevertheless, it seems that two kinds of inferences - the statistical one from sample to population and the substantive one from the empirical data to the nature of the phenomenon - are often confounded (Bakan, 1966, 1969), so that lower p values are interpreted as stronger evidence of the validity of a theoretical proposition. (See Bakan, 1966 for evidence of the use of this criterion for publication decisions.)

It is sometimes stated that the significance test is a necessary prerequisite for judging whether or not a certain result is worth interpreting (Winch & Campbell, 1969). A "statistically significant" result (i.e. p at least $\leq .05$) does not in itself, however, seem to be either necessary or a sufficient basis for making a decision whether the data are worth interpreting or not. First, if the expected result is not obtained (p higher than .05) the investigator or anyone who has any belief in the theory, will probably spend a good deal of time trying to understand and explain the outcome, instead of automatically rejecting the hypothesis. Judging from the "Discussion" parts of published reports this is a very common reaction, and certainly a recommendable practice if one is concerned with theory development. Second, if the result is significant this does not necessarily indicate that the data are worth interpreting, since the null hypothesis can be assumed a priori to be false anyway (Rozeboom, 1960; Bakan, 1966). However, in most laboratory experiments the number of observations is rather small, and if the obtained p value is clearly significant, the test result <u>as such</u> may be taken as a sufficient indication that the observed difference is worth theoretical explanation.

Although intended as a procedure for decision-making, the significance tests have limited applicability in this respect as regards basic research, mainly because investigators do not and cannot treat "acceptance" or "rejection" of a hypothesis as though these were decisions to be made. "Acceptance or rejection of a hypothesis is a cognitive process, a degree of believing or disbelieving which, if rational, is not a matter of choice but determined solely by how likely it is, given the evidence, that the hypothesis is true". (Rozeboom, 1960, p. 423, italics in original.)

An entirely different thing is that the computation of p values, confidence intervals, strength of relationships, or variance components serve as a very useful basis for the investigator in his attempts to interpret empirical data and further his theory. They form one part of the total evidence available, and their numerical values certainly are of importance as a guide to theoretical thinking. But using predetermined p values as a criterion or decision rule for how to think about and try to understand a phenomenon seems to constitute a severe restriction on theory development. It has also been argued (Rozeboom, 1960) that published research should include, besides the information "significant" or "not significant" at a specified level, the actual t- or F-values or confidence intervals and other descriptive statistics which may facilitate the reader's forming his own opinion about the meaning of the results.

This is not to say, naturally, that experiments should not be planned and carried out in such a way that the probability of chance results is minimized. From the point of view of validity it is a basic necessity to rule out the rival hypothesis that results are due to chance. What they mean, whether they are worth explaining, and whether they support the theory is largely determined by the researcher's knowledge, ingenuity, and involvement in the problem under investigation.

II:3c. Analysis of change

Studies of communication and attitude change by definition imply that the proper measure of treatment effect is a change score in the dependent variable.

Many, if not most, experiments on communication and attitude change have used a before-after design, and in most cases the analysis has been based on raw change scores, formed by subtracting pretest from posttest scores. Such raw change scores have since long been shown to be unreliable, primarily because they are systematically related to random error in the measurement of the dependent variable (McNemar, 1958). In recent years a number of techniques for estimating true change scores have been suggested, and in a review of the various methods Cronbach & Furby (1970) conclude that in cases where changes are consequences of experimental treatments there is neither need for nor advantage in using measures of change at all. Instead, they recommend, as has been done by other authors before (McNemar, 1958; Lord, 1963), the application of analysis of covariance with the pretest and/or any other suitable variable as covariate.

The common use of raw change scores is likely to be a threat to the validity of experiments on communication and attitude change, but no published reanalyses of data are available which could indicate how serious this problem is. Some of the earlier, "classic" studies (Hovland et al, 1949; Hovland et al, 1953) contained no analysis of individual scores, but compared the proportions of subjects changing in expected and opposite directions. Apart from the crude nature of the data, such an analysis is applicable of course only in simple experimental designs.

II:3d. Conclusion

In research on communication and attitude change it is difficult to avoid reactive measures unless the conceptualization of attitude is more explicitly geared to relationships between cognitive-affective structures and overt behavior. Attitude change is primarily conceived of in terms of cognitive-affective dynamics, and theoretical development has been directed towards more refined models of "mental structures" rather than identification and classification of behavior reflecting such structures (cf. Deutscher, 1966). In lack of a theoretical basis for connecting attitudes with a variety of overt behavior, most researchers in attitude change use more or less direct measures of a self-report nature, which of course may be highly reactive.

The reactivity of measures employed in an experiment, however, is a function of the nature of the measurement procedure in interaction with characteristics of the experimental setting. A completely non-reactive measure requires that subjects are unaware both of their participation in the experiment and of the fact that their behavior is observed. In research where self-reports or similar verbal responses are assumed to be the best or only available indicator of the phenomenon under study (like in much research on communication and attitude change), there seems to be a need for arranging experimental settings so that subjects are not aware of their participation in an experiment. In most cases this probably means leaving the laboratory in an opportunistic search for situations which are relevant to the theoretical problem at hand and which to some extent can be controlled by the investigator. Although such situations may be rare in the field of communication and attitude change, it seems that there is no other solution to the problem of reactivity in measurement of attitude change. Theoretical development of the relationships between on the one hand various classes of behavior and on the other hand an underlying cognitive-affective structure may open the road to other indicators of attitude change which are non-obtrusive or at least less obtrusive than traditional measures.

II:4. THE RESEARCH SETTING

Previous sections have primarily been concerned with various characteristics of laboratory experiments which more or less clearly have been demonstrated as possible threats to internal or external validity. In the following pages emphasis is more on ecological validity, i.e. to what extent the variables, relations, and processes studied in the laboratory are representative of naturally existing situations. At this point the discussion cannot be made in general terms, and the interest is focused on experiments on communication effects, especially effects in the form of attitude change. The underlying assumption is that a relevant area of application for such research is mass communication situations, but it is not implied that results of particular experiments were ever intended to be generalizable to such situations. The question is to what extent experiments on communication and attitude change in general are carried out in such a way that the research area as a whole can be assumed to result in theories which are ecologically valid in the field of mass communication.

First, however, some comments on ecological validity are required.

II:4a. The laboratory and the "real world"

Mundane realism. A rather common criticism of laboratory research which seems to be based on doubts about ecological validity,

is that the research setting is "unrealistic". As pointed out by Aronson & Carlsmith(1968) in their discussion of experimentation in social psychology, this claim for realism may mean different things. Laboratory experiments are bound to be unrealistic in the sense that they are atypical of everyday realities - from most people's point of view the laboratory setting is obviously not very familiar. It is evident that investigators usually do not intend that the experiments be realistic in this sense. Some authors have argued that experiments are in fact to be conceived of as models or analogues of the real world (Chapanis, 1963; Wärneryd, 1970), but it has also been stated that "in general an experiment should not be regarded as a model of the real world" (Björkman, 1970, p. 95). Both standpoints are reasonable, depending on what is referred to by the expression "real world". It can clearly not be required that an experiment be an analogy to a certain identifiable external situation which in some sense is representative of the happenings in the outside world. Rather, it is a model of particular psychological processes which are assumed to occur in a variety of real life situations. The question of fidelity or the realism of the experiment concerns whether the experimentally arranged situation actually activates those processes and whether they are "real" in the sense that subjects are involved and concerned. Such experimental realism is an essential part of the attempts to achieve internal validity, but it should not be confused with "mundane" realism (Aronson & Carlsmith, 1968) which refers to the similarity between the laboratory setting and a real world situation.

Experimental design and procedure are a result of theoretical considerations and, although there is always uncertainty about the validity of the operationalization of variables and processes under study, mundane realism in the laboratory automatically increases neither internal nor ecological validity. In fact, it may have quite the opposite effect, simply because the imitated "real" situation may take on quite a different meaning to the subjects in the laboratory, or because the special characteristics of the laboratory will be even more conspicuous, increasing suspicions or role-related behavior among subjects.

Any experiment must be based on some kind of theory, however inarticulate or non-scientific that theory may be. The theory essentially consists of assumptions about real-life phenomena, based on more or less systematic observations but also on imagination and intuition. The assumptions refer to the existence of relationships among a set of selected concepts and possibly to the nature of these relationships. The important word here is selected, which indicates that the theory can encompass only a part of the real-life phenomena, and in the behavioral sciences usually a very small part. To the extent that the theory encompasses relations and interactions between variables as they occur in the natural habitat of the phenomena in question, the theory may be said to have ecological validity.

An experiment focuses on a particular set of variables contained in the theory, and through the experimental procedure these variables and the nature of their relationships are given empirical meaning. Depending on the nature of the theory, the stage of theory development, and the goals of the investigator the experimental situation may require high or low mundane realism (cf. Weick, 1969). Whichever the case, the results of the experiment can be interpreted only on the basis of a theory, and the crucial question is whether the theory can be judged to have ecological validity (cf. Björkman, 1970, and p. 8 above). If so, all research which can justly be interpreted in terms of that theory may also be said to have ecological validity. In other words, the ecological validity of an experiment must be judged on the basis of both the nature of the guiding theory and the value of the experiment as a test of that theory (internal validity). In this perspective the degree of mundane realism in the particular experiment becomes wholly irrelevant.

Isolated processes. This leads to another aspect of the artificial nature of the laboratory setting, viz. the high degree of simplification and isolation of the phenomena under study. Although this is a basic requirement for controlled experiments, not to take the interrelatedness of variables and processes into

account may have serious consequences in terms of ecological validity. Again, this is not a fault of the laboratory experiment, but of the theories determining what should be isolated and simplified. Ecological validity of research in a field requires that theory development is not only a result of experimentation but of search for variables which are ecologically relevant. Such a search means going back to the problem from which the research originally emanated, be it practical or theoretical (Wärneryd, 1970), and extending attention to the ecological characteristics of behavior which are relevant to that problem. It is necessary to "feed real-world notions into psychological theories and models" (Björkman, 1970, p.97). It seems that much experimentation in psychology takes place rather in a closed system where new variables are introduced mainly on the basis of previous experimental results.

A related aspect of lacking realism in laboratory research is that the phenomena studied may be psychologically or socially insignificant. Similar criticism has been directed towards sociological field research, and this issue obviously has more to do with the interests of investigators and the characteristics of their theories than with the methodological approach. It does not seem reasonable to assume that the laboratory experiment as such precludes the study of interrelated processes which are psychologically or socially significant.

The issue of realism in laboratory experiments is thus not so much a question of similarity between experimental and natural settings, but it rather concerns the validity and relevance of the theories and models underlying the experimental procedure, and the capability of this procedure to activate the processes studied. To the extent that the latter is considered impossible to achieve in a laboratory setting, naturalistic experiments may be preferable, because they can encompass the more complex phenomena covered by ecologically valid theories.

II:4b. The communication situation in the laboratory

Isolated processes. The most conspicuous "isolated process" aspect of laboratory research on communication effects, which has often been pointed out, is the forced exposure to experimental messages. Exposure under such conditions is of course wholly alien to most natural mass communication situations, and the problem is further enhanced by the probably highly restricted variation in attention level in the laboratory. Research on selective exposure (Freedman & Sears, 1965a; Abelson et al, 1968), uses and gratifications in media consumption (Lundberg & Hultén, 1967; McQuail, 1969), and communication as a transaction process (Bauer, 1964) shows that the receiver may have chosen to expose himself and attend to a certain kind of information for a number of different reasons and with various expectations. It seems very probable that the impact of a message is dependent on the goals and motives associated with the act of exposing oneself to the message. In fact, message impact in the context of mass communication must be conceptualized as a result of the message and the act of exposure in combination. Research certainly should try to disentangle the two components and study the relationship between them, but that requires a different approach than has been used so far. There has been a considerable amount of research on exposure behavior and its determinants as well as on attitude change, but nothing on the crucial question of the relation between them. For example, would a person be more or less influenced by a message if he had chosen to expose himself because of its practical utility rather than its reassuring nature? (Cf. Canon, 1964; Freedman, 1965a.) Or, how would the presently rather inconsistent hypotheses about the effectiveness of fear appeals (Higbee, 1969) look, had the various experiments included self-selected exposure?

Another typical aspect of communication effect studies is the common practice of measuring the dependent variable immediately after exposure. To study immediate responses is of course quite appropriate if the theory is concerned with short-term effects,

but the question is how immediate? Weick (1967) has used the term "the confrontive laboratory" to indicate, among other things, that the subject (receiver) is forced to take a stand without much chance to reflect on the issue and the message received, or to postpone decision. In research on the effect of a forewarning message, preceding the actual persuasive attempt, Freedman & Sears (1965b) found that it made a difference whether the forewarning was presented two or twenty minutes before the experimental message. Even such short time periods may give the subject the opportunity to process more fully the information received, thus reducing the risk that the investigator obtains superficial, fleeting responses and possibly making the results less affected by compliance. The claim is not that short-term, direct effects necessarily are irrelevant or ecologically invalid, but that the immediate response measures may reflect processes which the theory is not concerned with (internal validity).

Related to the time perspective of effects is the social isolation of the subject in the laboratory. Again, there is a considerable amount of research and theorizing on the interaction and communication behavior within groups, but laboratory experimentation on persuasion has not attempted to integrate direct responses to communication stimuli with subsequent communicative behavior in a social setting. The isolated subject responding immediately has no opportunity to maintain cognitive inconsistency awaiting further information, or to validate his responses aginst those of other persons. After all, talking to other persons is a most typical human behavior, and presenting material for interpersonal communication is an important aspect of mass communication. It should be possible to study two so closely related phenomena in conjunction, not only in the survey type of research so far carried out by sociologists, but also with an experimental approach in the laboratory. Clearly, laboratory studies would still be limited to short-term effects, but it would be a step towards ecological validity. Some laboratory studies have been carried out where delayed measures of effect were utilized (e.g. Kelman & Hovland, 1953; Kelman, 1958; Freedman, 1965b), but theoretical considerations

about long-term effects have been of a mainly intraindividual character, not including aspects of social interaction.

It should be emphasized again that the problem does not lie in the laboratory approach as such. The solution is not to leave the laboratory entirely, even if "tempered naturalness" (Weick, 1967) may often be a promising alternative. From a theoretical point of view the goal is not simply to learn about, say, attitude change as a result of a complex process including selective exposure, individual information processing, and interpersonal communication, but rather to understand the nature of that complex process and how its characteristics are related to attitude change. The relative importance of the components in the process and their interdependence cannot be studied properly without a fair degree of control and systematic variation, which is usually impossible in a natural setting. It may require a wholly different conceptualization of what is a laboratory and of the experimental procedure, but for the purpose of theory development the "natural experiment" is no universal solution.

The size of effects. It has often been stated that considerable attitude change is a rare thing in natural mass communication situations, and it is assumed that forced exposure and lack of social interaction among subjects in the laboratory tend to exaggerate the amount of attitude change (Hovland, 1959). This may be true, but more important with regard to generalizability, however, is that the difference between treatments may be smaller or larger, or even change direction, depending on the presence of forced exposure and social interaction. In a comparison between, say, two ways of presenting the arguments there is no reason to believe that the effect of forced exposure is equally strong in both conditions. Similarly, the nature of any interpersonal communication subsequent to the exposure is very likely affected by the way in which the arguments were presented.

Another factor that makes for larger effects in the laboratory as compared with those in field settings is the nature of the issues

on which information is presented. Whereas field studies have typically centered on rather major issues, like political attitudes of preferences (cf. Lazarsfeld et al, 1948; Berelson et al, 1954) it is evident that in much laboratory research subjects confront trivial issues, issues in which the latitude of noncommitment is extensive, issues that are not ego-involving (Sherif et al, 1965). The reason of course is that the experimenter must select topics towards which attitudes can be expected to change as a result of rather limited information input.

The concept of ego-involvement is rather vague, and its relationship with susceptibility to persuasion has been a matter of controversy for some time. The social judgment approach to attitude change (Sherif et al, 1965) and the related "own categories" operationalization of ego-involvement has convincingly demonstrated, however, the importance of the relationship between involvement and change. But even if the applicability of most experimental results and theory on communication and attitude change is limited to non-involving issues, this is not necessarily a serious threat to ecological validity in the context of mass communication. In fact, there is reason to believe that much material presented in the mass media is exactly of this nature as viewed by large segments of the audience. Only minor parts of the information transmitted by the mass media are of high personal importance or direct relevance for everyday life. Indifference and lack of knowledge are common phenomena, and "issues with high ego-involvement may be less prominent in daily interaction than we suppose" (Weick, 1967, p. 65). Important effects of mass communication may be due to the fact that the audience is non-involved and indifferent to a wide range of issues (latitudes of noncommitment are extensive, Sherif et al, 1965).

II:4c. Conclusion

It is evident that most laboratory research on communication and attitude change has not been guided by a theoretical framework which can be assumed to have ecological validity with respect to mass communication effects. Experiments have been based on various theories related to scattered aspects of the mass communication process without taking into account some of the important relationships known or assumed to occur in natural mass communication situations. The reason, of course, is that there is no social-psychological theory which is specifically concerned with mass communication and attitude change (apart from certain attempts to formulate general propositions at a fairly "grand" level; e.g. Klapper, 1960). Instead, theoretical development tends to take place inside the traditional categories of psychology - phenomena are compartmentalized and studied in terms of perception, cognitive processes, social interaction, etc. Since, in real-life situations, these various aspects of behavior are interrelated, theoretical gaps exist which inhibit both conceptualization and empirical research as regards certain classes of real-life problems.

It seems that attempts to develop ecologically valid theories on mass communication and attitude change require that concepts and relationships are distinguished more on the basis of an overall conceptualization of the individual and social processes involved as well as the situational contexts in which they occur, rather than on the basis of traditional research areas. It is probable that a variety of other relationships than have so far received attention would emerge and be considered important. For example, it would seem difficult to maintain the present distinction between the uses and gratifications approach and the so-called effects approach. The relationship between kind of gratification derived from the message and susceptibility to influence would become an essential problem, as would the relationship between message characteristics and the probability of interpersonal communication following exposure.

One can argue that it is necessary to start with very restricted theories which may later be integrated into complex conceptual systems with a higher degree of ecological validity. This may be true, but there is a real risk that such integration never takes place - restricted theories thrive and develop in isolation from each other and take on independent scientific status. Integration requires going back to relevant real-life situations - a step which has so far not been taken in social-psychological research on mass communication and attitude change.

II:5. INDEPENDENT VARIABLES

Stimulus variables in experiments on communication effects usually relate to characteristics of source, message, or channel. Interpretation of results on attitude change typically requires consideration of organismic variables like initial attitude or familiarity of the issue. The major part of theoretical formulations define communication stimuli in terms of the affective-cognitive structure of the receiver (cf. the review by McGuire, 1968) and also include various organismic variables interacting with or operating independently of the stimulus variables. This emphasis on the organism is a natural consequence of the phenomenon under study (attitude change), but it has contributed to a lack of interest in classification of stimulus variables, and systematic analysis of message content is rather rare.

II:5a. The S-O-R-R paradigm

The stimulus variables (the message) in communication experiments, are ordinarily quite complex, and the experimental manipulation often refers to variables which are very difficult to relate unequivocally to characteristics of the stimuli. By defining the independent variables in terms of the receiver's perception of the message, it is possible to transform and reduce multidimensional and hard-to-describe variations in the message to variations in a

single psychological variable. The many studies of the effectiveness of fear appeals provide illustrative examples. The underlying model there follows an S-O-R_1-R_2 paradigm, where the dependent variable (R_2, attitude change or action) is supposed to be related to or dependent on the perception of the fear-arousing nature of the message (R_1, fear arousal), which in turn is related to the characteristics of the message (S).

The theoretically independent variable is fear arousal, but the methods of varying this variable have consisted of simultaneous variation in an unknown number of physical, syntactical semantic, stylistic, and content dimensions of the message. Such stimulus variations are typically described only summarily in the published reports, and the experimental conditions are instead defined through measures of audience perception of the message (cf. Nowak et al, 1966; Higbee, 1969). In some cases attempts have been made to distinguish various dimensions in the message which are varied orthogonally (cf. Nowak, 1966), but these dimensions or the resulting message structure are still defined in terms of audience perception.

This way of defining stimulus variables is of course very common in several fields of psychology, but it does offer some problems of practical as well as theoretical nature. From a theoretical point of view it may lead to some uncertainty about what is actually studied - which is the independent variable "explaining" the attitude change: is it message form or is it audience perception of messages?

Studies on fear arousal and attitude change have shown a good deal of inconsistent results, but it seems impossible to judge on the basis of published reports whether the inconsistency refers to the relationship between S (message form) and R_1 (fear arousal) or between R_1 and R_2 (attitude change) or both. A comparison of results from the various experiments requires that either the messages or the fear arousal variable can be described in the same terms, but none seems possible. It would have been much more fruitful to start off by developing a standardized measure of R_1, and then study its relationship with R_2 using a number of methods to vary R_1.

If a consistent relationship were found, there would be reason to study the (causal) relationship between S (message characteristics) and R_1. The confusion in one and the same experiment between variations in S and variations in R_1 is theoretically inhibiting. It offers two alternative ways of analyzing data having different meaning, and one of which may lead to erroneous conclusions (the internal analysis based on classification of subjects according to perceived fear). The absence of a standardized measure of R_1 furthermore results in difficulties in interpretation of inconsistent results - what in one study is classified as "high fear arousal" may in another study be called "low fear". This is further complicated by the frequent use of only two or, at most, three levels of fear arousal.

From a more practical point of view message variables defined in terms of audience perceptions offer a similar problem. If someone wants to replicate an experiment, or if a professional communicator tries to apply research results, a rather unequivocal description of the stimulus materials used in previous studies is required. This is the more important as the measure of R_1 is often associated with a high degree of uncertainty - if this measure lacks reliability or validity, it becomes difficult in replication or application to check the extent to which the stimulus variations chosen actually did result in the intended variations in R_1. In studies of fear arousal and attitude change the common practice has been to use quite crude methods for measuring R_1 - usually a few rating scales or questions (cf. Higbee, 1969). The presence of inconsistent findings is not surprising in view of these circumstances, and the same line of reasoning is applicable to other theoretical formulations in research on communication and attitude change (e.g. the issue of discrepancy /cf. McGuire, 1966/, or ego-involvement /cf. Sherif & Sherif, 1967/).

II:5b. Irrelevant stimulus properties

The complexity of stimulus materials and the lack of systematic analysis of content not only make replication and application difficult, but also aggravate the problem of irrelevant stimulus properties being offered in explanation of results obtained. The investigator has certain a priori ideas about how message form and content will affect the receiver's perception (R_1), but these stimulus properties are not part of his theory. He is free to choose any means available to create a certain R_1, and there seems to be a real risk that the researcher more or less unconsciously may vary the message in subtle ways which are not reported or do not follow from the description of the stimulus materials. A more appropriate approach would be to have the investigator himself simply state in as concrete terms as possible what his intentions are as regards R_1 and thereafter let someone else produce the stimulus materials (preferably in several versions which are all used). That would force the investigator to make explicit his expectations about what stimulus properties affect R_1, and the risk for subtle but systematic variations which are theoretically irrelevant but tend to support the research hypothesis about the relationship between R_1 and R_2 would be reduced. From an ecological validity point of view it would be useful if stimulus materials were produced or at least conceptually outlined by professional mass media communicators in accordance with theoretical specifications given by the researcher.

This leads to the question of representativeness of independent variables, which is one of the central points in Brunswik's (1956) claim for ecological validity. It is evident that theoretically oriented psychological research on communication effects has been very little concerned with representativeness of variables and variable values. Again, this is probably because of the strong emphasis on organismic variables which has made the problem seem irrelevant. Also, it may be rather difficult to define populations of "messages" or communication situations which are theoretically important. Representative design as proposed by Brunswik (1956)

poses certain questions as to methodology and analysis (cf. Sjöberg, 1971b), but even when using systematic designs the results would gain ecological validity from considerations of characteristics of the independent variables in natural settings. A basic point is of course that examination of relevant natural situations may lead to the introduction of new independent variables (cf. p. 38), but in an area where stimulus materials are very complex the operationalization of variables and selection of variable values are equally important in terms of ecological validity.
More interest in the analysis of mass media content and the use of professional communicators in the production of stimulus materials would seem essential in attempts to increase ecological validity in research on communication and attitude change.

II:5c. Conclusion

The reason for concern about independent variables in research on communication and attitude has to do primarily with the disinterest in stimulus properties, especially as regards message characteristics. To achieve ecological validity in the context of mass communication effects, other sources and criteria than the investigator's own predispositions will have to be utilized for the identification, selection, and operationalization of stimulus variables. One such source would be actual mass media content, analyzed in psychologically relevant categories. This may require cooperation with researchers in other fields and possibly also adaption of accepted methods of content analysis. Another source would be the professional mass media communicators and the underlying "theories" about receiver responses which guide their encoding behavior. The professional communicator acts very much like experimenter himself - he designs his message on the basis of (possibly implicit) hypotheses about S-R or S-O-R relationships. Attempts to make such professional "theories" explicit may reveal dimensions for describing messages which are not readily available to the social psychologist.

Another aspect of stimulus variables is the heavy reliance on more or less openly persuasive communications in most laboratory experiments. The search for ecologically relevant stimulus variables should not be confined to persuasive media content - the major part of the material presented by the mass media is not of a persuasive nature, but it is nevertheless (or exactly for that reason) of interest for research on mass communication effects and attitude change.

II:6. DEPENDENT VARIABLES

II:6a. What is attitude change?

Since the 1930's attitude has been a central concept in social psychology, and in the last twenty years theoretical interest in the field has flourished. Ostrom (1968), in a review of recent research, reports no less than 34 different approaches to attitude formation and change since 1950. Even before that time, of course, there existed a vast literature on the issue of attitude measurement (see Scott, 1968), and several major contributions as to the nature of attitudes as a significant social psychological concept (e.g. Thurstone, 1931; Allport, 1935; Newcomb, 1943; Sherif & Cantril, 1947). The more recent theoretical developments are associated with the trend towards laboratory experimentation in social psychology, which opened up possibilities to study more detailed aspects of attitude formation and change.

In a sense,the laboratory approach represents a somewhat different approach to the attitude concept. The sophisticated psychometric presentations of methods of attitude measurement, for example, stand in sharp contrast to the rather crude methods typically used in laboratory experiments. Likewise, there is a strong discrepancy between the broad concept of social attitudes as developed 30 - 40 years ago and the evaluative ratings of non-significant stimuli presently made by the subjects in many experiments on attitude change. "From studies labeled research on attitude and

attitude change, we might conclude that attitude is a blanket term covering any old judgment or opinion that the individual renders" (Sherif & Sherif, 1967, p. 111). Or, as stated by Allen (1966b, p. 283): "One of the shortcomings of traditional research in the attitude area is the excessive preoccupation with changing a response on an isolated (and apparently randomly selected) issue in the laboratory at the expense of research on the nature of attitudes."

It seems that this criticism is mainly directed towards the use of the term attitude to signify evaluative predispositions which are judged irrelevant for the study of social attitudes as a broad theoretical concept. As pointed out above, the preoccupation with trivial issues is not necessarily an indication of low ecological validity in the context of mass communication effects, but there is no doubt a considerable degree of confusion as to whether attitude change is a proper name for such effects. The majority of the theoretical approaches in recent years (cf. Abelson et al, 1968; Greenwald et al, 1968) are more generally concerned with the functioning of cognitive-affective structures and/or the acquisition of evaluative predispositions, and they should not be confused with the study of socially significant attitudes. This does not mean, naturally, that they are irrelevant for the latter purpose.

II:6b. Measures of attitude change

Apart from the reliability and (construct) validity of attitude measures typically used in laboratory research, and apart from the reactive nature of rating scales (cf. p. 28), the rigid adherence to one single kind of operationalization is in itself a problem of external and ecological validity. Although everyone agrees that attitudes have to be inferred from behavior, very few attempts have been made to enumerate or classify behavior assumed to

be relevant as indicators of attitudes and to suggest how such behavior can be observed in the laboratory (cf. Fishbein, 1966; the corresponding problem in psychology as a whole is discussed in Barker, 1963). The point made here does not refer to the relationship between verbal responses and overt behavior (see below), but simply to the necessity of developing attitude theories on the basis of a variety of operational definitions of attitude. A minimum requirement seems to be that different forms of verbal responses are observed. For example, most attitude theories assume that attitude towards an issue is related to perception of information on that issue. Rather than letting subjects agree or disagree with short statements the experimenter could present whole "messages" to be judged according to criteria which were selected on the basis of theoretical considerations about the relationship between attitude and perception. Such stimulus judgments of content and form of information would constitute a different experimental task than the attitude measures commonly used in experiments, and the observed behavior may have more ecological validity in the context of mass communication effects - one important consequence of attitude change may be that responses to subsequent information are affected.

Even if research on communication and attitude change is conceived of as the study of cognitive-affective processes without any consideration of the relationship between these processes and social behavior, it is clearly desirable that several methods are used for making inferences about cognitive-affective functioning. Likewise, interest should be directed not only towards attitude change in terms of evaluative responses - changes in other dimensions of attitudes, like complexity, salience, and differentiation, are equally important for a theory concerned with the influence of communication on cognitive-affective processes. Such studies are strongly underrepresented in the field of attitude change research. Although processes of attitude change as a consequence of communication may be of theoretical interest quite irrespective of the relationship between attitudes and social behavior, ecological validity requires a broader spectrum of methods of measurement. The question

of verbal vs. other behavior is possibly of less importance in this respect than the range of methods applied and to what extent they are of a non-reactive nature.

II:6c. Attitude change and behavior

A much more common concern with the validity of attitude change research refers to the relationship between changes in attitude scores and (changes in) overt behavior, as observed in some other situation, more or less distant in time and space, and more or less similar to the attitude testing situation. Cohen (1964) stated the issue as follows: "... attitudes are always seen as precursors of behavior, as determinants of how a person will actually behave in his daily affairs. ... Until experimental research demonstrates that attitude change has consequences for subsequent behavior, we cannot be certain that our procedures for inducing change do anything more than cause cognitive realignments; perhaps we cannot even be certain that the concept of attitude has critical significance for psychology" (p. 138).

In a review, Festinger (1964) concluded that there was practically no research geared towards this question, and that what was done did not give much support to a hypothesis about a positive relationship between attitude change and subsequent behavior. More recent studies have not produced consistent results, and it is evident that the problem offers several conceptual and methodological difficulties (cf. Ölander, 1969). The somewhat simpler question of the relationship between existing (verbal) attitudes and overt behavior has been subject to a fairly large amount of research, but the outcome is certainly theoretically disappointing (Wicker, 1969; Thomas, 1971). The relationship and time-order between changes in the two kinds of observables is less documented - there is theoretical as well as some empirical support for the occurrence

of attitude change subsequent to behavior change, but not the other way round (e.g. Brehm & Cohen, 1962; Zajonc, 1968).

The issue of communication, attitude change, and overt behavior can be approached from two different points of view. On the one hand, one may ask under what conditions an observed change in attitude score is related to the occurrence of one or more behavioral acts, and try to clarify the nature of such a relationship. On the other hand, one may ask under what conditions communication intended to affect attitudes is related to overt behavior. In the first case, attitude is regarded as a state of the organism, and the question is whether a change in this state will affect subsequent behavior. This may depend, among other things, on the permanency of the change and the relative importance of this state (attitude) as a determinant of behavior in a certain situation. In the second case attitude change is considered a cognitive-affective process mediating the relationship between communication stimuli and the occurrence of a particular overt behavior. The question then is not whether a change in attitudinal state affects behavior, but rather concerns the effectiveness of a particular mode of acquiring a behavioral predisposition (cf. Campbell, 1963; Nowak, 1968). Theoretical interest is focused on the relationship between content, form, and context of communication stimuli and the particular overt behavior under study, as mediated by cognitive-affective processes. It is not an organismic state but a psychological process that is seen as a determinant of subsequent behavior. The nature of this process and its relationship with overt behavior can reasonably be assumed to depend on the similarity between communication stimuli and the behavior in question, including the specificity of stimulus materials in pointing out that behavior.

Research based on theories about behavioral effects of communication - primarily the representational learning approach (e.g. Bandura, 1965) - has shown no interest in attitude change as a mediating process. In this kind of research communication stimuli are used to influence particular acts or act sequences, essentially by giving the receiver vicarious experience of those acts (and often their consequences).

Whether this results in attitude change, i.e. changes in verbal evaluations of the acts or objects involved in them, is not part of the theory. A mediating cognitive process may be postulated (cf. Bandura, 1965), but no attempts are made to observe this process or its result independent of the dependent variable.

Research on representational learning indicates that an attitude change approach to message form and content is less effective when the purpose is to affect a certain behavior in a certain situation (cf. Zimbardo & Ebbesen, 1969), but the point is that studies of attitude change typically do not have such a purpose. Very seldom have these studies had any theoretical basis for assuming that attitude change would be related to some particular subsequent behavior. Since behavioral variables have not been part of the theory, they have not been studied. Investigators may intentionally confine their interest to mental processes only, but an underlying assumption seems to be that attitude change is considered a significant social phenomenon in the sense that it is related not to particular behavioral acts but to a wide variety of behaviors in a wide variety of situations. The question whether communication intended to influence a particular attitude is an effective means to affect a particular overt behavior becomes irrelevant in that perspective, and this is possibly a reason for the widespread disinterest in the attitude change-behavior change relationship. On the other hand, the lack of theoretical concern over which classes of behavior are attitude-relevant and what situational or other variables affect the relationship is a severe shortcoming of attitude change research.

The attempts made so far to demonstrate empirically a relationship (or lack of it) between on the one hand a change in the attitude to a particular object and on the other hand particular behavior, conveniently observable in an easily available situation, are of little theoretical value. It seems evident that the controversy over attitudes and overt behavior is meaningless until there are some explicit theoretical linkages between overt behavior in social situations and attitude as a state of the organism or attitude change as a psychological process.

Some attempts in this direction have been made (cf. the review by Wicker, 1968), but on a fairly superficial level. An interesting

theoretical approach not directly related to persuasion situations has been taken by Weick (1966). In another context (Weick, 1967) he has pointed to the fact that the subjects in persuasion experiments are required to be very passive, not given any opportunity to engage in attitude-relevant behavior. This has bearing on the need for multiple indicators of attitudes as well as on the question of directing or motivating effects of attitude change on behavior. In addition, a change in experimental procedure in this respect might throw some light on the role of behavior as a means to stabilize any attitude change resulting from persuasive communications. The relationship between attitude or attitude change and overt behavior may be conceived of in many ways, and the ecological validity of research on communication and attitude change must be considered low until this relationship has been given more precise theoretical formulations.

II:6d. Conclusion

The points raised above lead to the conclusion that there is a need for theoretical development of the attitude concept. This is somewhat paradoxical in view of the tremendous amount of space in journals and textbooks, devoted to attitude formation and change, and to attitude measurement. Still,attitude is a very loose concept and there is little systematic analysis of the extent to which attitudes constitute an important aspect of various kinds of social behavior. This applies both to the more traditional idea of attitudes as enduring predispositions of the organism with respect to socially significant phenomena and to the experimentalist conceptualization of attitudes as more or less easily modifiable evaluative beliefs about or affective responses to practically anything.

Behavioral correlates of attitudes or attitude change have been so neglected in basic research and theorizing that the question is barely mentioned in the most comprehensive recent review of the area presently

available (McGuire, 1968). It seems that the significance of attitudes in terms of gross behavior has been considered a problem for applied research only, e.g. in the application of attitude theory to advertising , propaganda, etc. This is clearly a restricted perspective - even if interest is focused on attitude as a general social psychological construct or attitude change as cognitive-affective process, it is necessary to relate attitudes to a variety of observable behaviors or to other constructs. Due to the lack of such theoretical formulations most research on communication and attitude change has come up with rather stereotype operationalizations of the dependent variable. It seems that the theoretical concern over what attitude change <u>is</u> ought to be succeeded by a corresponding concern over what attitude change <u>means</u> with respect to other psychological processes and various classes of behavior.

II:7. ALTERNATIVE EXPERIMENTAL SETTINGS

The basic threat to the validity of psychological laboratory experiments is their social nature - the experimenter-subject relationship, the interaction between them, and the role-related behavior on the part of subjects. Improving the laboratory experiment must involve attempts to control for and/or assess the effects of such factors, and suggestions in this direction have been made by various authors. At the same time it has been argued that the characteristics of the experimenter-subject relationship must be altered, partly for reasons of validity, but especially on the basis of ethical and humanitarian considerations. Such arguments are closely related to criticism of the common practice of deceiving subjects as to the real purpose of the study or the actual meaning of what is going on, but there is also considerable concern about psychologically harmful treatments. The social nature of the laboratory experiment thus has two aspects - one methodological and one ethical, but attempts to overcome one kind of problem may affect the other as well. The points raised below primarily refer to the question of validity.

II:7a. Changing the subject's role

One obvious means of assessing the effects of demand characteristics, suspicions, or other role-related behavior is the post-experimental interview. As was mentioned above (p. 17), such interviews are not

very satisfactory since subjects tend not to reveal their beliefs about and actual behavior in the experiment, due to a "pact of ignorance" (Orne, 1968) with the experimenter. It seems probable that subjects will not convey the relevant information unless they feel that they are engaged in a joint effort with the experimenter and that the post-experimental interview is really an important part of the experiment itself. This may be very difficult to achieve, at least with college subjects, simply because students have firm expectations as to the way experiments are carried out.

One way of trying to avoid this dilemma might be to give subjects full advance information as to the purpose of the study, as has been suggested by Argyris (1968). The problem, of course, is that subjects may still suspect that they are not being told the truth. Such experiences have been reported by Kelman (1968).

Full information to the subjects probably will have to be accompanied by other modifications intended to render the subjects a higher degree of control in the experimental situation, thus altering the nature of the experimenter-subject relationship. Such attempts to equalize information and power in the relationship is the basic idea in Argyris' (1968) proposals as to improvements of the laboratory experiment.

Giving the subjects full advance information about the experiment will of course introduce a different set of demand characteristics and different kinds of role-related behavior. Even though it would probably be easier than in the normal experiment to find out what subjects actually responded to and how they perceived the situation, one may fear that the error variance increases substantially. Kelman's (1968) proposal to let subjects role-play a real-life situation in the experiment may be less problematic in this respect, since instructions as to the subjects' role clearly have to be more directive. Role-playing has been used successfully in the sense that regular patterns of response have been obtained (cf. Rosenberg & Abelson, 1960; Kelman, 1968), and in spite of the fact that it is difficult to judge how such results should be interpreted, it is not necessarily more difficult than in an ordinary deception experiment. A theory

based on role-playing may be just as useful as one based on deception experiments. Any mode of conducting social psychological laboratory experiments will be intrinsically confounded with demand characteristics and role-related behavior, and there is so far no way of judging which kind of demands are most harmful from the validity point of view.

In fact, it does not seem possible to assess the effects of demand characteristics other than very roughly. Orne (1968) has suggested the use of "non-experiments" or preinquiries as a complement to the real experiment, with the purpose of testing to what extent the experimental results may be due to demand characteristics. The idea is to present the whole experimental procedure, materials, measures, etc., to a control group, but not let them actually go through the experiment. These subjects are then asked to respond, as if they had actually been taking part in the experiment described. The arrangement is comparable to a kind of role-playing, but participants play the role of experimental subjects rather than a "real-life" role. To the extent that the non-experiment and the experiment give the same results, these indicate that subjects in the real experiment may have simply guessed what was expected of them. A parallel to this approach is Bem's (1967) studies derived from a theory of attitudes as self-descriptions. He described the procedure of certain experiments to his subjects and asked them how they thought the subjects in the actual experiment responded. It turned out that his subjects predicted quite accurately certain results which supported some "non-obvious" hypotheses in dissonance theory.

The various attempts discussed above to overcome problems associated with the social nature of psychological experiments seem to be closing a circle, leading back to introspection as a source of empirical data and to the use of non-naive subjects. It is evident, though, that there is no universal way of handling the social interaction aspect of the laboratory experiment. It is true that the mere awareness of being in a context which is most probably perceived as intrinsically manipulative will affect the subjects' behavior in some way or another. But in many kinds of experiments this may not be an important

problem, since there is only a main effect of the laboratory situation as such, which does not influence the interpretation of results. As to experiments on communication and attitude change, however, the issue is important, because it is often quite plausible both that extraneous factors interact with independent variables, and that the behavior observed reflects something other than the psychological process under study. For such studies, therefore, it seems especially desirable to avoid the awareness problem, which may mean leaving the traditional laboratory situation altogether.

II:7b. Disguised experiments

The awareness of subjects may refer to several things. One is the awareness of participating in an experiment - the subject knows both that someone deliberately has initiated a certain set of events with the purpose of finding out their consequences, and that he himself is acting as a "guinea pig". At a lower level of awareness, the subject may know that he is participating in an experiment, but he does not recognize the actual treatment he is exposed to, and does not know what kind of observations will be made. This would correspond to some of the "second order deception" experiments (Kelman, 1968), in which subjects believe that they are acting as accomplices to the investigator or as experimenters themselves.

A still lower degree of awareness would be when subjects have not noticed anything unusual happening in their environment but are exposed to a measuring device which may alert them to the previous occurrence of a "treatment". Finally, there is the completely non-reactive situation where subjects are unaware of both measures, treatments and, of course, participation.

It is evident that the "disguised experiment" (Campbell, 1968) or the "tempered nature" approach (Weick, 1967) may offer serious ethical

problems. In experimentation without the consent of subjects, strong restrictions must be put on the nature of experimental treatments and observations, and the invasion of privacy issue becomes pertinent. Experiments on communication effects may not be the most difficult case, but even there doubts may be raised. For example, varying the degree of fear arousal in a film on smoking and lung cancer presented to exhibition visitors (Leventhal & Niles, 1964) may have quite harmful consequences to those exposed to the high fear condition. In other situations, the use of a control group means that some people do not receive certain information. Depending on the nature of that information such a procedure might be ethically irresponsible. In cases where the dependent variable can be characterized as socially undesirable effects of information, say, aggressive behavior, it may not be possible, in a natural context, to demonstrate such effects experimentally in a wholly confirmatory way. A field experiment varying the amount of, say, media violence in different areas could not reasonably increase such media output, but would have to restrict experimental treatments to various degrees of reduction of media violence.

Experimentation without subject awareness or consent thus will have to be confined to such treatments and procedures as can be considered normal in the context where the study takes place. This is of course essential also from a methodological point of view - the more the experimental treatments deviate from daily events, the higher the probability that results are influenced by subject awareness. Especially if observation of the dependent variable involves reactive measures (e.g. interviews), the non-obtrusiveness of the treatments may be a crucial factor.

In communication research some attempts have been made to carry out disguised experiments (cf. reviews by Campbell, 1968; Haskins, 1964), but they are not common in academic research. Economic, methodological, and practical circumstances probably explain this fact - laboratory experiments are usually less expensive and it is often quite difficult to arrange a natural situation where units of observation are randomly assigned to treatments (especially when units are individuals). Fur-

thermore, disguised experiments in a natural setting often involve the need for cooperation or assistance from some outside agency. Academic researchers may often lack such contacts, and it is often difficult to obtain the necessary cooperation simply because the investigator wants a high degree of control over stimulus materials.

Another important reason to hesitate about disguised experiments in natural situations is that they might pose more problems than they solve. The only advantage over the laboratory experiment is the lack of awareness on the part of subjects, and even this may not always be fully attained. Any other kind of artifacts still must be considered, and there may often be more plausible rival hypotheses in the probably less controlled disguised experiment than in the laboratory. In other words, disguised experiments are no panacea, but rather a complementary approach which in some instances, where awareness is believed to be of crucial importance, is a worth-while effort. The more general advantage of disguised experiments in a natural setting is related to ecological rather than internal validity - they can encompass more complex phenomena and avoid the "isolated process" problem (cf. p.39). They may also force the investigator to utilize more representative experimental conditions. As mentioned in Section I:2b, this does not mean that results from naturalistic experiments are necessarily more generalizable - on the contrary they probably suffer from lack of specification as to the conditions under which results were obtained. In other words, even though they can potentially contribute to an ecologically valid theory, experiments in a natural setting have no inherent advantage as to external validity.

The final conclusion, then, is that psychological laboratory experiments are associated with certain unavoidable threats to validity, which are due to the social interaction between the experimenter and his subjects. In certain kinds of experiments these threats do not seem to be very serious, but in studies on communication effects there is a strong need to try modified laboratory procedures or to avoid altogether the subject's awareness of being in an experiment. The rest of this report presents an attempt to carry out a study where

a small part of mass media output was experimentally manipulated and where subjects were unaware of the existence of the experiment as well as of their own participation in it. Apart from that, certain considerations as to ecological validity guided the planning and the execution of the study.

III. A FIELD EXPERIMENT ON THE EFFECTS OF MESSAGE FORM

III:1. CONCEPTUALIZATION OF THE RESEARCH PROBLEM

The previous discussion of difficulties involved in the laboratory approach to mass communication effects indicates the nature of the considerations that led the present author to seek an opportunity to do experimental studies in a natural setting. Academic researchers generally have no access to the media for research purposes and, apart from studies on advertising or propaganda, field experiments with mass media content are very rare, especially as regards the electronic media. Due to certain circumstances, however, the present investigator was given the opportunity to exert some control over the production of a short TV program, to have it produced in two versions and have these broadcast simultaneously to different parts of the country. In other words, it was possible actually to manipulate communication stimuli experimentally in a natural context, where subjects were completely unaware of the experiment.

In spite of the methodological and theoretical difficulties involved, and in spite of the one-shot character of the experiment, the temptation to make use of this rather unusual opportunity was irresistible. Leaving the laboratory creates limitations on action and uncertainty in interpretation, and the criteria of "good research" are more or less unattainable. While the difficulties involved may have been underestimated when the study was decided upon, the decision was based on the conviction that development of a research area depends not only on neat sets of data but also on painstaking attempts to try new approaches.

The purpose of the experiment was to study the effects of message form on learning and attitude change. The message was a TV-program

of an informative nature, intended to be included in the normal program output at the Swedish Broadcasting Corporation. In other words, it was neither an educational nor a propaganda program. The topic treated was dental health in Sweden.

The basic research problem was of the same kind as that in the classic laboratory experiments on the effects of message form conducted by Hovland and his colleagues (Hovland et al, 1953; Hovland, 1957) or in those previously conducted by the present investigator on fear arousal and attitude change (Nowak et al, 1966). But the experimental treatments were both designed and presented in a naturally occurring situation, which had certain consequences for the selection and operationalization of variables.

III:1a. The approach to operationalization of independent variables

In a previous section (p.45) it was mentioned that studies of communication and attitude change often apply an S-O-R-R model, where complex variations in S are expressed in terms of receiver perceptions of the message (R_1). As long as the experimenter himself can produce and select the stimulus materials, his own preconceptions of the relationship between various message dimensions and receiver perceptions will guide the operationalization of experimental treatments. In a normal mass communication situation, however, the production of stimulus materials requires the cooperation of professional communicators - authors, photographers, producers, journalists, etc. These persons have their own ideas about the relationship between message characteristics and audience perceptions, and they have professional criteria as to which variations can and may be done in the experimental materials. In addition, several other restrictions on message form are introduced, mainly due to media policies and scarce resources in terms of time, personnel and money. The investigator thus has both the problem to define and communicate

to others what message dimensions he wants to vary experimentally, and to make others operationalize these conditions.

Following an S-O-R-R model, the investigator can avoid these problems and at the same time make the experiment as naturalistic as possible (at least as regards the stimulus materials) by allowing the production and design of the experimental messages to be guided by the conditions and criteria normally controlling the production of messages in the medium involved in the study. The experimenter indicates to the professional communicator what he wants to achieve by the different treatments in terms of audience perceptions (R_1), and he indicates how he thinks that these may be related to the dependent variables (R_2).

The experimental treatments are thus operationalized in accordance with the preconceptions of the professional communicator rather than those of the experimenter, but the latter decides upon the purposes of the treatments. These purposes are based on theoretical considerations while the means to reach them are determined by existing practice in the context in which the experiment takes place. From a theoretical point of view, this approach is no less satisfactory than that typical of most laboratory studies, where stimulus variables are defined in terms of subject perceptions. In both cases the experimenter must check to what extent the intended experimental treatments have been realized, and in both cases there is uncertainty as to which stimulus variables actually were studied. The difference is that message form becomes more "natural" or "normal" when the experimenter himself does not control the operationalization of the treatments. This, of course, is wholly in line with the idea behind a field experiment.

III:1b. Design of the experimental treatments

The field experiment was preceded by a laboratory experiment using the same type of TV-program, which should be explained somewhat.

Following up the author's earlier investigations of fear-arousing appeals, this experiment involved variations both in "fear" and "specificity of the recommendation" (see Fröjd & Olsson, 1968). These variations were performed independent of one another (orthogonal design). Thus, the dependent variables were chosen on theoretical grounds. It proved quite difficult, however, for the producer of the programs to find useful operational definitions which allowed great enough variation in the creation of the program. Given those (for that matter entirely normal) restrictions posed by technical, temporal and economic resources, as well as the demand of making "good television" worthy of being broadcast, it was extremely difficult to perform those variations corresponding to the theoretical specifications. No quite unequivocally definable message dimensions (e.g. the amount of "fear-arousing" pictorial material) could be varied enough without the program changing character in other respects. Other material must be filled in in order to satisfy time requirements or to make the program "good television", etc.

Difficulties of this type naturally occur even in normal laboratory experiments, where the experimenter has complete control, but they were particularly exaggerated here. One of the conclusions drawn as a result of the laboratory experiment was that if the message (i.e. the TV-program) is to be created for use in a normal mass media context, the dimensions to be studied must be chosen on the basis of what a professional communicator considers to be meaningful and realistic alternative courses of action. It is necessary to consider his frame of reference, that is, his professional "theory" of the relation between the design of the message and audience reactions. Such a theory must underlie all those decisions made by the communicator in the process of producing his message, decisions which form a complex hierarchy, from the choice of over-all approaches to the subject matter, down to details such as the choice of words or camera angles (cf. Nowak & Wärneryd, 1969).

The question then becomes, on what level of this decision hierarchy is it reasonable to choose independent variables for a field experiment of the present type. Variables on a high level, of an extremely

general nature, are often difficult to define in concrete terms, let alone to unequivocally define in operational terms. Variables on a very low level, details in the production of the message, cannot be expected to achieve any observable effects, because the message itself is so complex that the relative importance of minor details cannot be expected to be very great.

The conditions demanded of the field experiment were (1) that the design of the programs should differ substantially and (2) that the producer should perceive the versions of the program as realistic alternatives. In addition, two basic conditions must be filled: first, that for economic reasons the program material from the laboratory experiment must be utilized, and secondly, that the program versions used in the field experiment must contain the same factual information. The latter demand had ethical as well as methodological origins: the variations were to concern format variables only, and furthermore, it would be impossible for the Swedish Broadcasting Corporation to broadcast different programs to different parts of the country if they did not at least contain the same information.

The choice of program designs was made on the basis of the already extant programs which had been used in the laboratory experiment. These were reduced to a single version, and the experimenter together with the producer developed descriptors which both could accept as meaningful to characterize this version. The program was described in terms of both stimulus variables and expected receiver reactions. A corresponding description was then made of a hypothetical program which could be said to represent a clearly different, but from the producer's point of view an entirely realistic way of conferring the same information. Generally speaking, this entailed choosing descriptors which posed opposites or definite contrasts to the characteristics of the existing program.

A full description of the text and pictorial material of the two experimental programs is given in Nowak (1971a) and in the separate Appendix in Swedish to this report. The original program was characterized as serious, somewhat heavy, quite fear-arousing, systematic

but hardly stimulating for anyone uninterested in the subject, rather anonymous or impersonal, having the character of a lecture. The text of the program comprised interviews with a number of experts intermixed with a speaker text. The speaker never appeared on camera; rather, the speaker text was accompanied by illustrations of the various stages of dental diseases, both schematic and realistic, or by other appropriate illustrative material. The interviews were presented primarily in synch, that is, with the speaker appearing on camera. No special effects or "tricks" were used to increase the viewer's involvement or attention. A couple of brief musical interludes marked the passage between different segments of the program.

One might say that the design of this program assumes that the public should be interested enough in the subject matter to remain in front of the TV-set solely for the information conferred. In the case of the alternative program, the producer proceeded from the assumption that the public would be only slightly interested in the content, and the aim was, by means of the program design, to capture the viewer's attention and increase his inclination to remain in front of the TV-set.

Consequently, the alternative program was intended to be "light" and somewhat humorous in tone, somewhat personal and having some entertainment or "thrill" value in addition to the factual content. Contrasts with the existing program were achieved largely as follows: a dramatic story-line was added, with a female TV-journalist as the main character. The program described how this reporter made a program about dental care and dental diseases, how she studied various types of material and interviewed specialists and "the man on the street". The dental information content was thus presented as a sort of "film within a film". Now and then this scheme was broken entirely, and the main character/TV-journalist addressed herself directly to the audience with advice or recommendations. A few humorous or absurd scenes were also included.

The interviews with specialists were the same in both versions, but in the "lighter" version the synchronous portions were shortened and replaced with other pictorial material.

The two versions of the program thus contained largely the same factual information, but the context in which the information was presented and the general character of the versions was different. As mentioned above, the independent variable was defined in terms of intended audience perceptions of the program, and the primary purpose was to relate this variable to learning and attitude change with respect to dental health. Audience perceptions or other receiver responses cannot, however, be related to specific stimulus variables, since the program versions varied in a number of dimensions. Still, it was judged both necessary and desirable to investigate to what extent quite different sets of stimuli (message forms), defined in terms of audience perceptions, can be expected to result in at all observable differences as regards learning and attitude change in a natural mass communication context.

III:1c. Dependent variables

Although the primary dependent variables are learning and attitude change, it is evident that program form may affect these variables in several ways, and it is therefore necessary to include other variables, which may be viewed both as dependent and as control variables to be held constant.

To study the effects of program form on learning and attitude change it is of course necessary that the groups compared have had the same chance to be influenced by the program, e.g. they must have been exposed to the program to the same extent. The number of persons in each group who happen to watch TV when the program starts is not dependent on the form of the program, but the program format must be assumed to affect the length of exposure, i.e. the proportion of the audience that stays in front of the TV through the whole program. While length of exposure must be held constant when one wants to compare groups as regards learning and attitude change, it is also an important dependent variable.

Even groups which have seen equally much of the program may differ as to the nature of the exposure situation. A TV-viewer may very well be occupied with other simultaneous activities, and his perception and processing of the information presented may be interfered with in various ways. One may discuss the program with other persons viewing it, which, while sometimes increasing interest and attention, sometimes acts as a distraction. Conditions of this kind must be investigated so that they may be held constant in the comparison of learning and attitude change. But again, this type of data may also form a dependent variable, affected by the characteristics of the program.

III:2. HYPOTHESES AND PROCEDURE

III:2a. Hypotheses

A natural mass media exposure situation may produce stronger or weaker effects than those obtained in a laboratory situation. On the one hand, external factors may inhibit reception and processing of information presented as well as tend to reduce or eliminate differential effects of small variations in program form. Relationships demonstrated in the laboratory may not be generalizable due to the noise level in a natural situation.

On the other hand, the receiver's freedom of action in the natural situation may increase the differential effect of variations in program form. The demand characteristics in the laboratory and the captivity of the audience serve to suppress differences in attractiveness between the experimental programs. The natural situation not only permits self-selection in terms of actual exposure to the experimental treatment, but also allows attention, motivation and social interaction to vary, free of experimentally induced limitations. The present experiment was primarily intended to avoid this type of limitations of the laboratory situation.

In view of the theoretical vagueness and the methodological uncertainties associated with the experiment, no formal hypotheses were formulated. Underlying the study, of course, were certain expectations as to the differences in effect between experimental treatments in the

dependent variables. These expectations were based on the assumption that such audience perceptions as were intended by the "lighter" version would be related to higher levels of attention, motivation, and involvement. In terms of the uses-and-gratifications approach to mass communication research, this version was designed to be more immediately rewarding (Schramm, 1954) than the other version. Thus, it was hypothesized that the lighter version would keep a larger proportion of viewers through the whole program, and that, with length of exposure constant, it would also be more effective in influencing knowledge and attitudes. In other words, it was expected that the lighter version would be more effective in two ways: by keeping a larger proportion of the potential audience, and by making a stronger impact in influencing knowledge and attitudes among those actually exposed to the program.

III:2b. Design of the experiment

The experimental design of the investigation was determined largely by practical circumstances concerning the opportunities of utilizing the various broadcasting transmitters. It was clear from the beginning that the only plan the technical staff of the Swedish Broadcasting Corporation would accept as technically feasible would be to utilize one local transmitter for the one version, while the other version would be broadcast as usual via the other transmitters. It would, of course, have been desirable from a methodological point of view to distribute the program versions over several transmitters, but this was impossible for both technical and economic reasons (several film copies would have been necessary).

Even the choice of the local transmitter to be used was determined by the capacity available at the time of the broadcast. It could be predicted, however, that either of two areas (Norrköping or Örebro) would probably be the locale. The transmitter used in the experiment was

the one in Örebro. One of the program versions was thus broadcast in the Örebro area; since it was the heavier, lecture-type version, it will be called the L-version. The other version was broadcast over the rest of the country. As it was the more entertainment-oriented, it will be referred to as the E-version.

The programs were almost to the second the same length, 15 minutes. None, other than a few persons within the Swedish Broadcasting Corporation and within the research team, knew that two versions of the program existed, and even fewer knew that one of the versions would be broadcast over only a single area. The separate broadcast began after the presentation of the program and ended before the closing announcement. In other words, there was no way of noticing that an experiment was being conducted unless one lived in the fringes of the Örebro district and could - and did - receive the program on more than one channel.

The program was entitled "Tooth by tooth" (Tand för tand) and was broadcast Monday, 8th September, 1969, at 8.05 PM. Immediately preceding the program came News and Weather, 7.30 and 8 PM, respectively, and immediately following came a program of modern ballet, called "Gay Time". No special advance publicity was given for the program, neither in the announcement nor in the program schedule made up by the Swedish Broadcasting Corporation, which is made available to the newspapers. Only the rather neutral sub-title, "A program about teeth and dental care" appeared there.

III:2c. Data collection

The data was collected on several different occasions using various methods. The following list indicates the available data:

1. Baseline study. This consisted of five factual questions about dental disease and dental care, asked in telephone interviews with circa 350 persons in a nation-wide survey during the period 4th-7th September, 1969. The survey was undertaken by the Department for Audience and Program Research (PUB) of the Swedish Broadcasting Corporation as an appendix to their daily measurements of audience size. (See Appendix 1.)

2. Incidental technique survey. A telephone interview survey was conducted on the evening the program was broadcast. The interviews commenced 5 minutes before the program ended and continued for roughly one hour. The sample consisted of circa 100 persons in Orebro and circa 100 persons in a comparable city (Linköping). (See Appendix 2.)

3. Mail questionnaire study. This was a postal survey conducted with a sample of 100 persons in Örebro and 100 persons in Linköping. One week prior to the broadcast each of the respondents received a letter in which the Swedish Broadcasting Corporation requested their participation in a study of how the TV-audience perceives various programs. They were asked to watch television during a one-hour period the following Monday and were also informed that they would receive a questionnaire in advance of the broadcast which they should answer immediately after having watched the programs indicated. All were asked to watch television between 8 and 9 PM, Monday, 8th September. The questionnaire, sent so as to arrive on the date of the broadcast, consisted primarily of a number of rating scales (Semantic differentials), to be used in evaluating the program "Tooth by tooth". (See Appendix 3.)

4. Producer study. This study treated TV-producers as experimental subjects. They viewed the two versions of "Tooth by tooth" and then filled out the same questionnaire as was used in the Mail questionnaire study. The ratings were done twice — once as an expression of the producer's own perception of the programs, once as an indication of his expectations as to how the audience would rate them. (See Appendix 4.)

5. Effects study. A telephone interview survey was conducted during the period 9th-11th September, using two different samples. The first was a sample representative of the Swedish population as a whole, comprising some 400 persons (E-sample), the other consisted of 300 persons in Örebro and its environs (the L-sample). While the questionnaire contained questions of the type asked in the Incidental technique survey and in the Mail questionnaire study, the primary content was made up of factual knowledge questions and questions concerning attitudes toward dental care and dental disease. (See Appendix 5.)

6. Information-seeking data. One week after the broadcast a letter from a national association promoting dental health (Tandvärnet) was sent to 300 persons. The letter contained certain information about the association and a list of its publications. The addressee was offered the publications of interest to him free of charge. The sample was taken from a sample of respondents interviewed by PUB about their TV viewing during the period 6th-8th September. Thus, it was possible to determine whether or not each of those respondents who ordered publications had seen the program "Tooth by tooth". (See Appendix 6.)

The studies discussed in the present report are primarily the Mail questionnaire and the Effects studies, but occasional references are made to the other studies. Technical details concerning the design and procedures followed in the studies may be found in the Appendices. The principal results of the Baseline study, the Incidental technique survey, the Producer study, and the Information-seeking study are also to be found there. An account of the different variables considered in the analysis of the Mail questionnaire and the Effects study is given in the presentation of results.

III:3. CHECK OF THE EXPERIMENTAL TREATMENTS

Before an analysis of the effects of the program versions can be made, it is necessary to check to what degree the intended experimental treatments, that is audience perceptions of the versions, have actually been achieved. Such a check can be made either by studying the same persons studied with respect to the dependent variables, i.e. the sample of the Effects study, or by studying other persons, who can be assumed to be representative of the Effects subjects. Both methods have been employed here, but the principal data concern persons other than the subjects of the Effects study.

III:3a. The method used

Audience perceptions had to be studied using a separate sample because of the restrictions placed upon the experimental procedure, especially the fact that only telephone interviews were possible. To obtain valid measures of audience perceptions of a television program via telephone interviews (which, furthermore, in some cases were carried out up to three days after exposure to the program) must be regarded as practically impossible. In this case the aim was to obtain a measure of the momentary perception of the program, which was expected to affect the effectiveness of the program in conferring information or influencing the receiver. Thus, the measure should be taken in close conjunction with exposure to the program, while the exposure should occur in a quite natural situation.

These demands are difficult to satisfy, and the course of action pursued here (Mail questionnaire study, see Appendix 3) can be regarded as only partly satisfactory. In the Mail questionnaire study the subjects received a letter asking them to watch certain programs in their homes at a specified time. They knew that they were expected to evaluate the programs, and they had access to the rating scales at the time of their exposure to the program. This is hardly a "natural" exposure situation, but it is, nevertheless, a more natural situation than if the measures were carried out in a laboratory setting. On the other hand, one has less control over how and when the evaluations are made. (It may be noted that only those persons who can have filled out their forms at the latest one day following the program have been included in the analysis; a few questionnaires returned later have been excluded.) Ideally, the only way of satisfactorily achieving such measures would be to conduct personal interviews immediately following the broadcast - an alternative which had to be excluded for practical reasons.

Since the Mail questionnaire study involved an entirely untested method, a number of follow-up telephone interviews were made in order to determine what degree of difficulty had been encountered in filling in the questionnaire forms. No such difficulties could be detected, however. All of the subjects (N = 45) felt that the instructions were easily understood, and nearly all (N = 41) found filling out the questionnaire a simple task. Two-thirds of the subjects were alone when they filled out the questionnaire, and of the remainder only four had discussed the program with anyone else. In one case the person spoken to had had a different opinion than the subject.

Possible sources of systematic error in the method used are of less importance, given that they affect the two versions of the program in the same way. It is hard to see any reason why this should not be the case, and it appears reasonable to assume that the results give a satisfactory picture of the <u>differences between the versions</u>, as regards the audience's perceptions of the program.

The composition of the samples, rate of non-response, etc. is presented in Appendix 3, along with the text of the cover letter, instructions and questionnaire.

III:3b. The choice of rating scales

The use of semantic differential scales for evaluations of the present type of stimulus material is quite common. Apart from the general theoretical arguments which recommend this technique (see for example Snider & Osgood, 1969), it was without question the most appealing alternative for the self-administered type of testing involved here. The choice of scales to be used was not as obvious, however. Many studies have shown that the results obtained by means of the semantic differential are quite dependent on the interaction between the scales and the object of evaluation (see, for example, Nordenstreng, 1969; Heise, 1969).

Two points of departure were chosen. PUB had previously carried out a relatively comprehensive survey with the aim of isolating a number of dimensions of the perception of TV-programs (Nilsson, 1970). Four factors were extracted, each defined by four semantic differential scales. It seemed only natural to include the sixteen scales in the Mail questionnaire study. In a later, more extensive study (Nilsson, 1971), similar factors were extracted, but one of them was split in two.

Due to the specific nature of the experimental programs and to the intended differences between them, however, additional scales had to be included. These could be chosen only on a subjective basis, with reference to the verbal descriptions made by the producer and the experimenter when planning the format of the two versions. The following scales were employed in the study:

Scales taken from PUB's study:

Factor I	Factor II
1. important - trivial	5. violent - peaceful
2. informative - uninformative	6. wild - tame
3. useful - useless	7. noisy - quiet
4. relevant - irrelevant	8. lively - placid

Factor III

9. serious - humorous
10. controversial - non-controversial
11. easy - difficult
12. light - heavy

Factor IV

13. engaging - dull
14. varied - monotonous
15. good - bad
16. attractive - repulsive

Additional scales

17. interesting - uninteresting
18. simple - complex
19. frightening - relieving
20. pleasant - unpleasant
21. unusual - common
22. personal - impersonal
23. entertaining - boring

III:3c. Intended differences between the program versions

In order to check the extent to which the experimental treatments have been achieved, one must define the differences intended in operational terms. This has already been discussed to some extent, but it is possible in terms of the scales presented above to sketch the pattern of differences expected if the aim of the experimental variations had been realized.

First, it may be noted that no differences were intended to be achieved as to ratings of the factual content of the programs - both versions were to contain the same amount and the same type of information. Thus, the scales for Factor I were not expected to show any differences, as they were intended primarily to express such a judgment of the programs. As for the other factors, it may be said that the E-version was produced with the aim of increasing its values

in Factor II (above all, a more rapid tempo) and to place it more toward the adjectives "entertaining", "personal", "pleasant", "relieving", "light" and "humorous". No differences were intended regarding the more generally evaluative scales, but the occurrence of such differences must be checked, as the over-all evaluation of the program may be of importance regarding its effect on the dependent variables.

III:3d. Results of the Mail questionnaire study

Diagrams III:3a and III:3b illustrate the profiles of the two versions in terms of the scales employed. As can be seen there, the intended differences between the versions have been largely achieved regarding the direction of differences, with the exception of the scales in Factor II, where only one scale shows a difference in the intended direction.

Any closer analysis of these profiles would not be too worth-while, since the reliability of the individual scales cannot be ascertained, and the two experimental groups differed slightly in composition. Thus, it has been necessary to take appropriate combinations of the scales and to make a more detailed comparison, holding constant those background variables which may have influenced the results.

The forming of indices. Initially, two factor analyses were performed on the material, taking the two groups together. (Principal axes, orthogonal rotation according to the varimax method; see Westling, 1967). In the first analysis, the number of factors was fixed in advance at four. The purpose of this procedure was to find out both whether or not the factors identified in the PUB study would turn up, and to what degree the additional scales could be assigned to these factors. The second analysis was open-ended with regard to the number of factors. It resulted in five factors, which explained

Diagram III:3a. Ratings of E- and L-versions (Persons born 1930 or later)

E = - - - -
L = ———

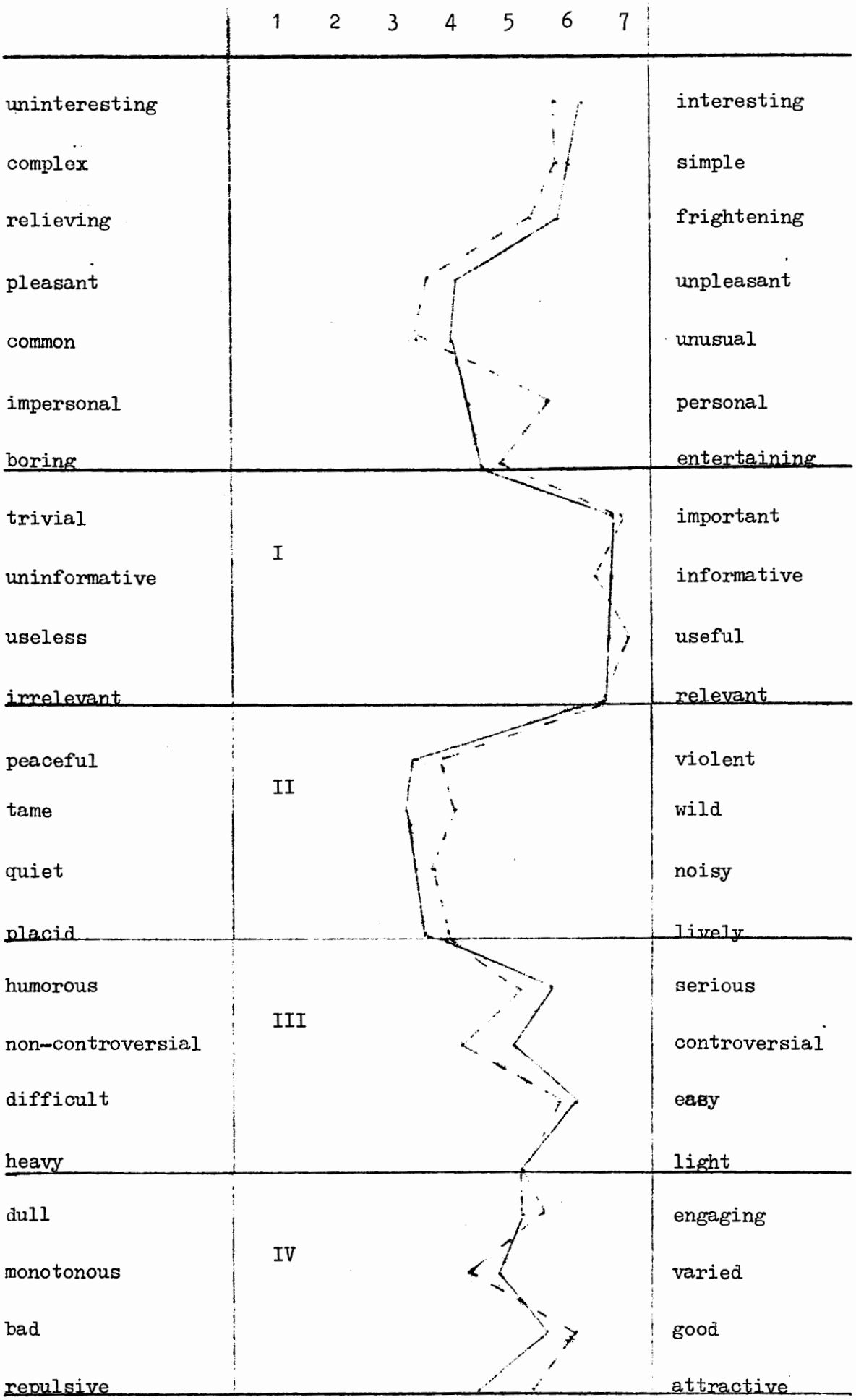
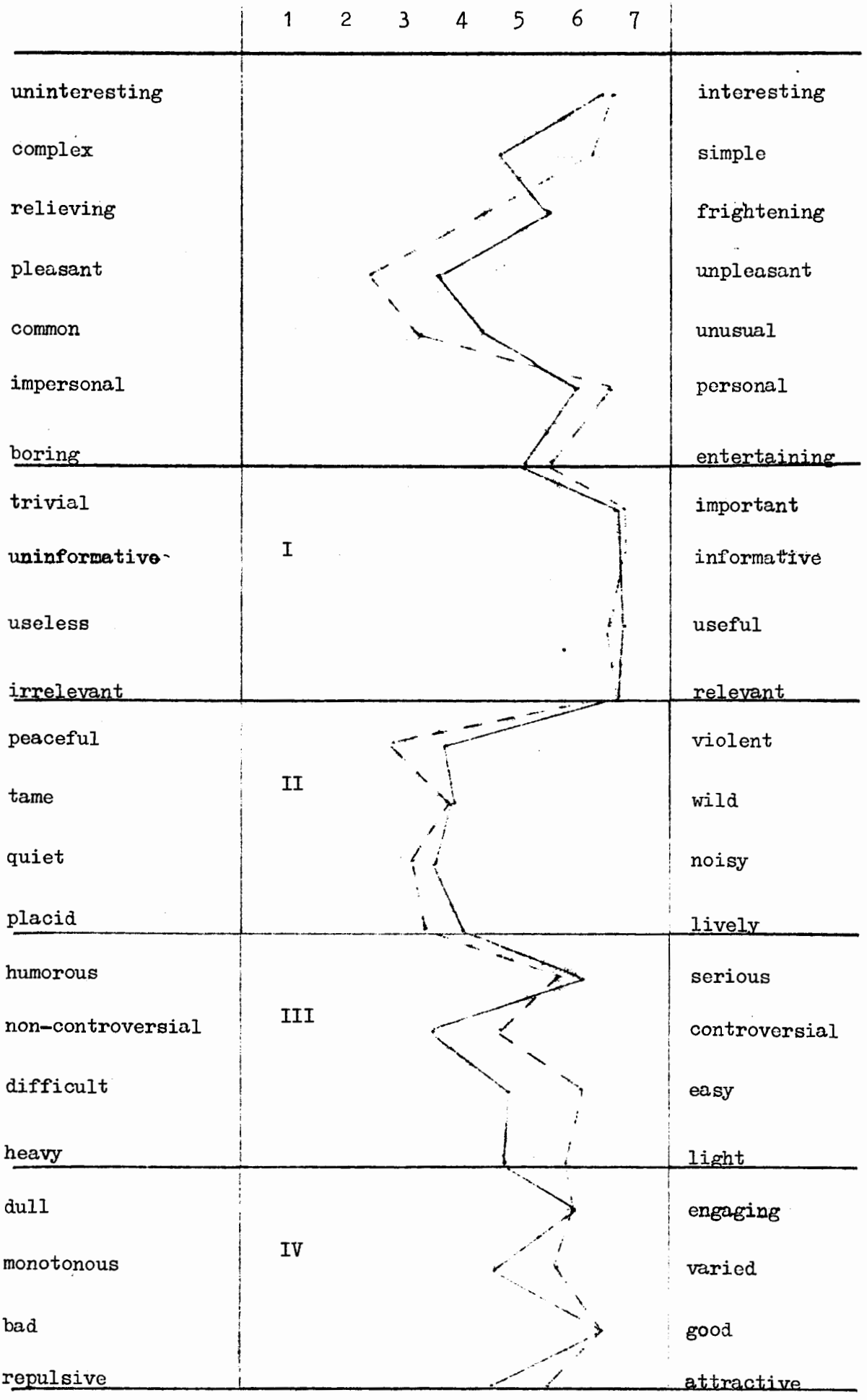


Diagram III:3b. Ratings of E- and L-versions (Persons born before 1930)

E = - - - -
L = ———



93 % of the variance. Neither of the analyses showed much agreement with the PUB study regarding Factors I and III, while the other two factors corresponded at least partly. Also, both factor solutions yielded quite "impure" factors - many scales had quite high loadings on two or more factors. Hence, neither of the analyses was accepted as sufficient basis for forming indices.

Instead, a cluster analysis was performed on the basis of the intercorrelation matrix and the five factor solution, with the aim of maximizing the intercorrelations in each cluster (McKennell, 1970). This resulted in seven clusters. The composition, average intercorrelation and internal consistency (alpha) were as follows:

Cluster	Scales	Average intercorr.	Alpha[1)]
Evaluation	Good Entertaining	0.63	0.77
Usefulness	Useful Relevant Informative	0.59	0.81
Involvement	Important Interesting Engaging	0.58	0.81
Comprehensibility	Simple Easy	0.55	0.71
Attractiveness	Attractive Light	0.53	0.69
Fear-arousal	Unpleasant Frightening Violent	0.46	0.62
Tempo	Wild Noisy Lively Unusual	0.32	0.65

These clusters are used in the following analysis.

Analysis of the indices. As can be seen in Appendix 3, the two groups of subjects in the Mail questionnaire study differed somewhat with respect to age and sex distribution. The group viewing the L-version

1) Alpha estimated according to McKennell (1970).

is dominated by Older persons and Low-educated, while those viewing the E-version were practically evenly distributed as to Age and Education (when these variables were dichotomized). Consequently, it was necessary to check to what degree these variables co-vary with the program judgments, both generally (main effect) and in interaction with the design of the program. If one combines the two groups, it becomes evident that the clusters Evaluation, Comprehensibility and Fear-arousal reveal a tendency toward main effect of Age and Education - Age co-varies positively with Evaluation and Fear-arousal, but negatively with Comprehensibility. The two latter clusters also co-vary positively with Education. In the following analysis, Age and Education have been held constant.

Tables III:3A and III:3B indicate means for the various clusters in different sub-groups. Focusing on those dimensions where clear differences between the versions occur, one finds that these are Involvement, Comprehensibility, Attractiveness and Fear-arousal. Further, the differences between the versions are statistically significant primarily in the Older group (those born 1930 or earlier) and in the Low-education group (elementary school only). The E-version was given higher values in Comprehensibility and Attractiveness by all groups, while it received consistently lower ratings in Fear-arousal. On the dimension Involvement, however, a tendency appears toward interaction with Age and Education. The E-version was rated higher among the Older and Low-education groups, while among the High-education group the opposite obtains, and among the younger subjects, there is no difference.

These results largely agree with the pattern intended to be achieved by the program variations, with the exception of the tendency toward interaction mentioned above. However, larger differences had, of course, been expected, which would provide a firmer statistical basis for judging the achievement of the experimental treatments. The only deviation from the intended pattern occurred on the dimension Tempo; there occurred no systematic differences, and the most sizable differences were in favor of the L-version.

Cluster	Elementary education only		More than elementary education	
	E-version (N = 24)	L-version (N = 24)	E-version (N = 25)	L-version (N = 14)
Evaluation	5.94	5.56	5.52	5.54
Usefulness	6.64	6.86	6.77	6.74
Involvement	4.91 x)	4.51	4.71	5.08
Comprehensibility	5.98	5.42	6.32 x)	5.54
Attractiveness	5.42 x)	4.77	5.48	5.00
Fear-arousal	3.81 xx)	4.62	4.43	4.81
Tempo	3.59	3.66	3.71	4.41

x) = $p < 0.10$ with t-test for the difference between the versions

xx) = $p < 0.05$ with t-test for the difference between the versions

Table III:3A. Means of the two program versions for the clusters of rating scales in the Mail questionnaire study. Education held constant.

Cluster	Younger (born 1930 -)		Older (born - 1930)	
	E-version (N = 25)	L-version (N = 15)	E-version (N = 26)	L-version (N = 25)
Evaluation	5.38	5.07	6.10	5.86
Usefulness	6.72	6.76	6.72	6.87
Involvement	4.57	4.57	5.04	4.81
Comprehensibility	6.18	6.27	6.19 x)	4.96
Attractiveness	5.36	4.90	5.60 x)	4.76
Fear-arousal	4.68	4.87	3.58 x)	4.71
Tempo	3.79	3.62	3.56	4.03

x) = $p < 0.05$ with t-test for the difference between versions

Table III:3B. Means of the two program versions for the clusters of rating scales in the Mail questionnaire survey. Age held constant.

These results indicate, however, that the possibilities of relating learning and attitudes to the audience perceptions of the versions are restricted to those results pertaining to the Older and Low-education groups. Among the Younger and High-education groups the present results indicate that little or no differences between the versions, in terms of effects on knowledge and attitudes, may be expected on the basis of the theoretical considerations underlying the study.

III:3e. Program ratings in the Effects study

Program ratings in the Effects study consisted of questions asking the respondent to say whether each of a number of adjectives was Very appropriate, Quite appropriate or Not at all appropriate as a descriptor of the program "Tooth by tooth". These adjectives were chosen to correspond as closely as possible to the scales employed in the Mail questionnaire study, but certain modifications were necessary. Tables III:3C and III:3D summarize the results of these questions, with Age and Education held constant.

In Tables III:3E and III:3F the adjectives have been placed in groups corresponding to the clusters used in the Mail questionnaire study, and the average response percentage has been calculated for each group. The pattern of differences between the versions is largely the same as that found in the Mail questionnaire study, apart from the fact that the E-version has a higher value in Tempo, as was intended. The differences between versions, however, are consistently small, which is hardly surprising considering the considerably less sensitive measure employed in the Effects study. Furthermore, the groups involved in the Effects study were more heterogeneous with respect to length of exposure to the program - the values in Tables III:3C - III:3F have been calculated for all those who saw at least five minutes of the program. The results, however, show no tendencies

	Elementary education		Higher education	
	E-version	L-version	E-version	L-version
	(N = 55)	(N = 30)	(N = 39)	(N = 35)
Informative	96(51)[x)]	100(57)	95(41)	100(66)
Important	100(85)	97(77)	100(92)	100(94)
Engaging	66	63	70	80
Lively	40	30	38	32
Humorous	64	47	82	89
Easily understood	100(75)	93(80)	100(85)	100(83)
Unpleasant	27	27	16	31
Frightening	47	53	46	60
Uninformative	4	13	-	3
Placid	78	90	92	89
Boring	4	20	6	6
Unusual	47	56	59	60
Impersonal	4	3	21	9
Serious	77	90	77	74

	Younger (born 1930 -)		Older (born - 1930)	
	E-version	L-version	E-version	L-version
	(N = 36)	(N = 26)	(N = 58)	(N = 39)
Informative	95(39)[x)]	100(54)	96(52)	100(67)
Important	100(93)	96(89)	100(87)	100(85)
Engaging	58	61	72	80
Lively	25	23	48	36
Humorous	50	34	59	43
Easily understood	100(75)	99(81)	98(79)	97(82)
Unpleasant	22	23	22	33
Frightening	56	73	42	46
Uninformative	0	4	0	3
Placid	34	46	59	69
Boring	5	15	3	10
Unusual	61	58	47	59
Impersonal	11	11	11	3
Serious	75	81	77	82

x) Figures within parentheses indicate the number answering "Very appropriate".

Tables III:3C and III:3D. Percentage of those who viewed the program longer than 5 minutes who answered "Quite appropriate" or "Very appropriate" for the adjectives used in the program ratings in the Effects study.

	Younger (born 1930 -)		Older (born - 1930)	
	E-version	L-version	E-version	L-version
	(N = 36)	(N = 26)	(N = 58)	(N = 39)
Usefulness	98	98	98	99
Involvement	79	79	86	90
Tempo	46	39	45	34
Fear-arousal	39	48	31	40
Attractiveness	38	27	36	31
Evaluation	95	85	97	90
Comprehensibility	100	99	98	97

Table III:3E. Program ratings for different age groups in the Effects study. Adjectives grouped according to the clusters of the Mail questionnaire study. Figures indicate average percentages for adjectives in clusters. (High percentage = high cluster value.)

	Elementary education		Higher education	
	E-version	L-version	E-version	L-version
	(N = 55)	(N = 30)	(N = 39)	(N = 35)
Usefulness	96	94	98	99
Involvement	83	80	85	90
Tempo	31	20	23	22
Fear-arousal	37	40	31	46
Attractiveness	44	29	53	58
Evaluation	96	80	94	94
Comprehensibility	100	93	100	100

Table III:3F. Program ratings for different education groups in the Effects study. Adjectives grouped according to the clusters of the Mail questionnaire study. Figures indicate average percentages for adjectives in clusters. (High percentage = high cluster value.)

Dimension	Own ratings		Expected audience ratings	
	E-version	L-version	E-version	L-version
Usefulness	5.72	5.70	5.63	5.54
Involvement	5.19	5.11	5.19	4.85
Evaluation	4.94	3.69	5.00	3.81
Comprehensibility	5.14	4.72	4.75	4.06
Attractiveness	4.89	3.44	4.17	2.92
Fear-arousal	3.52	4.61	4.32	4.96
Tempo	4.07	3.75	4.29	3.60

Table III:3G. The Producer study. Producers' ratings of the program versions. (The order of presentation was not completely balanced; see Appendix 4. Each order of presentation has been given equal weight in calculation of means.)

directly contradicting the results of the Mail questionnaire study. Bearing in mind the small number of observations on which the percentages are calculated, nothing indicates that the respondents in the Effects study perceived the program versions significantly differently than the subjects in the Mail questionnaire study.

The relatively small differences between the versions indicated in the viewers' ratings naturally reflect the fact that the producer and the experimenter overestimated the effect of the program format on audience perceptions. A similar overestimation was also revealed in the Producer study (see Appendix 4, or Nowak, 1971b, included in the Swedish appendix to this report) - the producers expected considerably greater differences in audience evaluations between the versions than actually occurred. Their expectations corresponded, however, to the pattern of differences aimed for in the field experiment (see Table III:3G above). Thus, the producer who, together with the experimenter, designed the program versions, cannot be said to be atypical regarding his expectations of audience reactions.

III:4. THE EFFECTS STUDY: PROBLEMS OF INTERNAL VALIDITY

The preceding analysis indicates that the experimental treatments would appear largely to have been achieved, although the differences between treatments are not very large. Before analyzing the results of the Effects study as regards the dependent variables, however, it is necessary to discuss the conditions for drawing conclusions as to the effects of the program versions, to check the internal validity of the study. The design of the experiment is such that this point is quite difficult to ascertain, a circumstance which, for that matter, is quite common as soon as one leaves the sheltered milieu of the laboratory.

As indicated in the description of the Effects study (Appendix 5), three circumstances complicate the interpretation of the results, that is to say, the judgment as to whether the program format has had any effect on the dependent variables. These are:

1. The experimental subjects were not individually assigned to the alternative experimental treatments (either randomly or via matching), but rather the groups were formed on the basis of geographical location. The independent variable "design of program" is thus entirely confounded with the demographic characteristic, "living in Örebro or not".

2. The subjects in each of the experimental groups exposed to the experimental treatment have themselves chosen to do so (unaware, however, that they were taking part in an experiment). Thus, the experimental treatment has been applied on the basis of self-selection.

3. Measurement of the dependent variables in the experimental groups has been undertaken only after the experimental treatment. That is, the study has an "after-only design".

III:4a. Comparing naturally existing groups

The first point is the most serious and in principle prohibits a definitive interpretation of the effects of the experimental treatment. No statistical or logical method can after the fact satisfactorily correct for initially existing, uncontrolled differences between groups formed by non-random procedures (Lord, 1967).

The basic idea in experimental design is to maximize the likelyhood that the differences between experimental groups and control groups which may be observed following an experimental manipulation are "due to" or can be "explained by" this manipulation. In the classical experimental situation this is rendered empirically probable by holding the situation under as strict control as possible, and it is made logically probable by randomly allocating the units of observation to experimental and control groups. In this way the groups can be expected on probability grounds to be initially like or comparable (prior to the experimental manipulation).

If, however, the experimental and control groups are made up of already existing groups, e.g. persons from different geographical areas, it will never be logically probable that they were initially comparable in all respects capable of affecting the interpretation of the results. This remains an entirely empirical question.

It is necessary in both cases to make an empirically based judgment as to the likelihood that the result can be accounted for by something other than the experimental treatment (cf. Blalock, 1964, p. 22 ff.). As to the initial comparability of groups, however, a decisive dif-

ference arises in that while the critic of the classical experiment must demonstrate that the statistically expected comparability of the groups is or can be assumed to be empirically false, the critic of an experiment employing naturally occurring groups need only suggest an alternative interpretation which implies that the groups have not been initially like. In the latter case, the experimenter must show each alternative interpretation to be untrue or at least improbable.

The present case involves an after-only design, the observations having been made only following the experimental manipulation. In such cases it is not possible to determine directly to what extent the groups initially differed with respect to the dependent variable. This can be done only with the help of variables which are believed or known to co-vary with the dependent variable, but which have not been affected by the experimental treatment. If the groups differ in such variables, a possible systematic error (bias) can be eliminated by in various ways holding the variables constant. A common technique to this end is the use of analysis of co-variance, which also leads to greater sensitivity or precision in comparison of groups, by reducing the error variance (Fennessey, 1968).

This technique in no way alters the logical conditions for causal interpretation of the results achieved. Whatever may have "caused" the result is still a question of likelihood, which must be judged with respect both to how the experiment has been carried out and on the basis of a causal model. With the aid of analysis of co-variance the effect of one or several co-variates may be eliminated, but the possibilities of drawing conclusions of a causal nature about any remaining relationship between independent and dependent variables is still a question as to how probable it is that the groups were comparable from the beginning in terms of other variables than those held constant. The statistical technique eliminates one alternative explanation of the results (Towers et al, 1962).

The fact that the controlled variables are responsible for a good part of the variance in the dependent variable does not in principle change the situation. Even if the control variables (co-variates) and the

independent variable together explain the entire variance in the dependent variable, no logical argument sustains the idea that the independent variable should be the most probable cause of the effects remaining after controlling for the co-variates.

Similarly, it is often assumed that if groups were acted upon so as to make them like in terms of the co-variate, then the experimental manipulations would give the result obtained in the analysis of co-variance. There is obviously no ground for such a claim - the analysis only eliminates the possibility that any effects in the dependent variable have been caused by the co-variate.

Whether possible rival explanations of the results received are eliminated by means of analysis of co-variance or through other methods of holding variables constant, problems of interpretation are identical. Logically, conclusions as to causality in experiments using naturally existing groups must be based on the assumption that the groups look as though they have been randomized with respect to all variables not controlled. How the groups would appear had they actually been randomized is impossible to say; rather, the experimenter must make a judgment on the basis of theoretical considerations as to which characteristics of the groups or of the experimental treatment may have systematically affected the observed changes in the dependent variable. On the basis of a theoretical model the experimenter may indicate which variables he considers important, and he must seek to show that the groups are similar on these points or take statistical measures to eliminate extraneous influences on the dependent variable. The better and more comprehensive theory he is equipped with and the more data he has gathered about the groups, the better his chances of arguing for a certain interpretation of the results.

It is perfectly clear that identifying all the important variables which it may be necessary to control or to correct for as well as gathering data on the groups for all these variables poses large problems. The important thing, though, is that even if one has a well-developed theory and has data on all the significant variables, the interpretation of the experimental effect must be based on an

assumption that the groups, after statistical controls, look like they would after randomization. Thus, interpretation of an experiment using naturally existing groups must be undertaken on a quite different basis than when the groups have been randomized, and such experiments require both more empirical data and a broader theoretical frame-work than does the classical experiment.

III:4b. Self-selection

Self-selection is a problem of internal validity only insofar as it interacts with experimental treatments. Voluntary participation on the part of subjects is thus no threat to internal validity if those subjects actually participating can be randomly assigned to treatments. When, as in the present study, comparisons are made between naturally existing, self-selected groups, however, there is always the question whether the same variables are related to self-selection and whether these relationships are equally strong in both groups. If not, there is an indication that self-selection may be a source of bias in the comparison.

This kind of analysis is of course possible only when data are available for non-participating subjects as well as for those participating. In that case, one can find out whether frequency of self-selection is similar in both samples, and whether the same set of variables account for self-selection. To the extent that such is the case and self-selection can be fairly well "explained" by available data, it becomes a less plausible rival hypothesis in interpreting the results obtained in the study.

The other side of this argument is that, in an after-only design, data on non-participants may be used to test the comparability of the self-selected groups in the dependent variable. If in both samples the same factors account for self-selection and these are related to the dependent

variable in the same way, there is a basis to assume that any difference in the dependent variable between the two non-participating groups did exist initially also between the self-selected groups. The question of finding determinants of self-selection is thus important for two reasons in an after-only design: both to assess whether an effect of self-selection on comparability is likely and, if not, to test the comparability through an analysis of the non-participating groups.

In the present study, self-selection refers not only to exposure to the experimental treatment but also to the length of exposure. This offers certain problems of analysis which will be discussed in some detail in connection with presentation of results.

III:4c. The present experiment

The variables which, parallel to the independent variables, may have affected the values of the dependent variables in the two groups can belong to two main categories: those pertaining to characteristics of the experimental subjects (O-variables or "attributes" in Blalock's, 1964, terminology) and those pertaining to the subjects' situation or environment (S-variables or "forcings"). Within each of these categories may be found, on the one hand, stable variables which can be expected to have had roughly the same effect even if the experiment had been done at some other time or in some other situation, and on the other hand, variables which assume importance under temporary conditions arising in the situation at hand. Further, one can distinguish both specific variables, directly related to the subject matter of the program, and general variables, whose effect would have been significant regardless of the program content.

		O-variables	S-variables
Stable	General	Personality, socio-econ. variables, habits and expectations regarding TV-viewing	Local health situation, or public health activity
	Specific	Knowledge, attitudes, interest in dental health and dental care	Availability of dentists, local dental care campaigns or other activity concerning dental care
Temporary	General	Frame of mind, tiredness	Distractions or factors enhancing attention
	Specific	Acute dental problems, date of last visit to the dentist	Advance information about the program

Figure III:4A. Examples of various types of confounding variables.

Figure III:4A present examples of various types of confounding variables. Of these, the stable, specific O-variables are considered the most important, both because theoretically they can be expected to exert considerable influence over reactions to the program and the effects of the information, and because they may be assumed to be affected by or related to a number of other stable variables, of both the S- and O-types. The major part of the analysis is devoted to the question of whether the experimental groups can be assumed to have been initially comparable with respect to one of the principal dependent variables, knowledge. Certain other checks are possible, however.

Since the experiment has an after-only design, the initial knowledge of the groups cannot be examined at first hand. As indicated in the foregoing discussion, two courses of action present themselves. The first is to ascertain whether and to what degree experimental group membership may be related to variables which co-vary with the dependent variables, but which cannot have been affected by the experimental treat-

ment (indicator variables). If so, these must be held constant in the analysis of results.

The other strategy is to study the determinants and extent of self-selection. Should the groups be similar in these respects, a certain support is provided for assuming that the relationship between the dependent variables and exposure is equally strong in both samples. In that case, any difference between the non-exposed groups in terms of the dependent variables may be taken to indicate differences between the exposed groups.

III:4d. Analysis of group comparability

Relationship between membership in experimental groups and variables indicating initial knowledge. The Baseline study on a separate sample prior to the broadcast included five of the factual knowledge questions included in the Effects study. The responses to these questions co-varied, as was expected, with Age and Education (see Appendix 1), but also with the responses to the question as to whether the respondent had watched the TV-News on the evening prior to the experimental broadcast.

The correlations with Knowledge were as follows: Age -.25, Education .19, and News-viewing .15.[1] The multiple correlation is R = .28, which means that Knowledge can only barely be estimated with the aid of these variables. The corresponding multiple correlation in the Effects study was R = .24 for the total E-sample and .28 for the L-sample. Even though the correlations of these variables with membership in the experimental groups is hardly a satisfactory indicator as to the comparability of the groups, they may indicate whether any of the variables should be held constant in the analysis of the results.

---

1) Age was coded using four categories. Education has been dichotomized - the Low-group comprising those who report having "no other education than elementary school", the High-group including the rest.

	The entire samples	Those having seen something of the program
Correlation between group membership and:		
Age	0.03	0.03
Education	0.06	0.13
News-viewing	0.02	0.04

Table III:4A. Correlation between group membership and variables indicating prior knowledge.

The correlation between group membership and the three indicator variables is presented in Table III:4A, both for the entire samples and for the self-selected experimental groups. (Exposure is defined as "having seen something of the program".) The correlation for the samples as a whole is very low, but the experimental groups differ somewhat on the Education variable. While the correlation between Education and group membership is significant only at the .10-level, it would nevertheless appear that Education should be held constant.

Determinants and extent of self-selection. As may be seen in Table III:4B, News-viewing correlates quite highly with exposure to the program. In both samples Age and Education show low correlations with Exposure if News-viewing is held constant. These variables, which are rather highly correlated with the dependent variables, thus have had no direct effect on self-selection.

Correlation between Exposure and:	E-sample	L-sample
	Raw correlations	
Age	.12	.22
Education	-.07	.04
News-viewing	.58	.65
	Partial correlations	
Age; News-viewing held constant	.09	.01
Education; News-viewing held constant	.04	.09
News-viewing; Age and Education held constant	.57	.63

Table III:4B. Correlation between Exposure (= having seen something of the program) and indicator variables.

These correlations give no cause to suspect that self-selection has operated differently in the two samples. It would also appear that the pattern of TV-viewing during the evening is quite similar in the two groups: among those who were at all exposed, roughly 40 per cent have watched both TV-News, the experimental program and the following program, while approximately 10 per cent have seen the experimental program alone (see Table III:4C). The Table also shows that exposure frequency is also roughly the same in the two samples - the difference revealed is not significant, and it becomes even smaller when one calculates exposure frequency on the basis of those who had the TV-set at all turned on that evening.

	E-sample (N = 354)	L-sample (N = 221)
Had had TV turned on during the evening	46 %	51 %
Had watched TV-News	36	39
Had watched something of "Tooth by tooth"	32	36
Of those who had seen something of "Tooth by tooth":	(N = 114)	(N = 81)
Watched TV-News, TBT and "Gay time"	39 %	38 %
Watched TV-News and TBT	40	44
Watched TBT and "Gay time"	10	9
Watched TBT only	11	9

Table III:4C. The pattern of television viewing. Effects study.

Comparison of the non-exposed groups. Since self-selection would appear to have operated similarly in the two samples, it is reasonable to compare the non-exposed groups with respect to the dependent variables. Table III:4D shows means and standard deviations. The category "Non-exposed" has been defined here as all those in the respective samples who have not seen any part of the experimental program.

The differences between the two samples are in all respects slight, and the analysis gives no reason to believe that the experimental groups have been dissimilar in terms of the dependent variables. The difference obtained in the attitude variable is far from significant.

	E-sample (N = 240)		L-sample (N = 140)	
Knowledge variable[x)]	M	s	M	s
Total Non-exposed	63	15	64	16
Elementary education	61	16	61	15
Higher education	65	15	66	15
Attitude variable[xx)]				
Total Non-exposed	120	33	117	33
Elementary education	117	35	113	34
Higher education	126	30	121	32

x) Index formed by multiplying (average number of correct answers to factual knowledge questions) x 100. Cf. Section III:6a.

xx) See Section III:6b for a description of the index.

Table III:4D. Means and standard deviations among Non-exposed in the two samples.

Other control variables. Certain other variables, similar to those presented in Figure III:4A, may also be mentioned. The presence of distracting factors, for example, may be closely related to both whether the subject viewed the program alone or together with someone else and whether or not he was engaged in some other activity while viewing the program. Questions on these points were included in the interview, and it was revealed that in both samples roughly three-fourths of the subjects viewed the program in the company of another person. A somewhat larger proportion of the subjects in the L-group than in the E-group reported that they watched the program and engaged in no other activity (e.g. ate/drank coffee, read, talked): L-group 85 %, E-group 69 %.

Another question dealt with the date of the subjects' most recent visit to the dentist; neither here did there appear any differences

between the averages of the two groups. There was also a question as to the subjects' perceived state of their teeth, but this is a dubious control question, as the responses may well have been influenced by the design of the program. Nevertheless, in both groups the most common response was that the subject thought his dental health was "neither better nor worse" than that of people in general.

It should be pointed out that some of the variables discussed above were included in the Incidental technique survey. The findings presented in Appendix 2 agree with those of the Effects study.

Finally, it should be mentioned that the questionnaire used in the Effects study included three factual knowledge questions which were intended as a control for the comparability of the groups. These dealt with certain aspects of dental hygiene and dental disease which were not taken up by the program. One question concerned the relationship between smoking and paradontitis, another the probability of paradontitis among pregnant women, and the third the utility of eating fresh fruit after meals as a means to prevent caries. It was not possible to evaluate these items in advance, and they turned out not to correlate at all with other knowledge questions.

However, an index of the control questions, formed in the same way as the knowledge variable in Table III:4D, shows fairly equal means for the exposed groups: in the Low-education group M = 68 for the L-sample as against M = 62 for the E-sample, and in the High-education group M = 67 for the L-sample as against M = 56 for the E-sample. These differences are far from significant, but considering the absence of correlation between the control variable index and the knowledge index it is doubtful whether this result can be interpreted as a failure to demonstrate non-comparability.

III:4e. The comparability of the experimental groups - a summary

The data obtained do not contradict the assumption that the groups were initially comparable in the dependent variables, but as was mentioned earlier, it is impossible to prove that this assumption is correct. The relatively high correlation between TV News-viewing and exposure to the program indicates, however, that characteristics of the subjects with regard to the subject matter of the program have had less influence on self-selection than general TV-viewing habits and simple situational factors. It is a well-known fact that the audience size of a TV-program is highly dependent on the position of the broadcast in the evening's program schedule and on how many persons viewed the immediately preceding program. (This relationship was particularly strong at the time of the experiment, as Sweden then had only one TV-channel.) Of those persons viewing the experimental program, only roughly 20 per cent in both samples had not seen the immediately preceding TV-News, and it is reasonable to assume that individual dispositions or interests can have had a greater influence on the likelihood of exposure primarily in these groups.

Thus, nothing suggests that the dependent variables may have had different importance for the likelihood of exposure in the two samples. The practically total agreement between the mean values of the Non-exposed groups thus supports the assumption that the experimental groups did not differ in the dependent variables. The other control variables, while of less importance, do not contradict the rest of the analysis. In the following analysis, the experimental groups will be assumed to have been comparable in terms of initial knowledge, and situational variables will be assumed to have had the same effect in each of the groups. Due to the correlation between education and group membership, however, education will be held constant. Since that variable has been dichotomized, the analysis is made separately for the two educational groups, rather than by means of analysis of co-variance.

III:5. RESULTS OF THE EFFECTS STUDY. EXPOSURE BEHAVIOR VARIABLES.

The first group of dependent variables concern the behavior of the viewers during their exposure to the program. As has been indicated above, the design of the program was expected to affect the length of the time of the exposure and to influence the degree of activity during the broadcast. The exposure time variable has been assigned one of the following values, which correspond to the response choices offered on the questionnaire:

0 have seen no part of "Tooth by tooth"
1 watched less than 1 minute
2 watched 2 - 3 minutes
3 watched 4 - 5 minutes
4 watched 5 - 10 minutes
5 watched 10 minutes or more (or the entire program)

The exposure activity variable was measured by two separate questions: one asking whether or not the respondent had discussed dental health or hygiene with anyone else who watched the program, and a question which asked the respondent to say whether he had followed the program "closely" or "less attentively".

III:5a. The exposure time variable

Quite a number of those who had seen something of the program did not

see the program in its entirety. A large majority (approximately 80 %) of the audiences of both versions saw more than 5 minutes (see Table III:5A). Only slightly more than half of the viewers of the L-version watched more than 10 minutes, however, compared to three-quarters of the E-group. This difference between the program versions remains when various categories of background variables are compared (Table III:5B) - in all subgroups a larger proportion of the L-group had watched less than 10 minutes. This result supports the hypothesis as to the relative attractiveness of the two versions. The difference is especially pronounced among those with more than an elementary school education. As the table indicates, the difference between the versions is significant for the group with higher education than elementary school, but not for the low education group.

As noted earlier, an individual's inclination to watch a given television program must be assumed to depend on his knowledge, attitudes and interests regarding the subject matter. The difference between the experimental groups as regards exposure time thus may arise from initial differences in these variables. While it is impossible to reject this possibility, one factor, apart from the analysis of the comparability of the groups (Section III:4), would indicate that such is not the case. If a viewer quits watching a program primarily due to lack of interest in the subject matter, it would seem reasonable that he quit watching toward the beginning of the program, that is, as soon as he has got a clear idea of what it is about. If, on the other hand, the viewer quits watching because he doesn't care for the program as such, i.e. the format and content together, it will probably have taken him some time to get an impression of the program. As was noted in Table III:5A, there was no difference between the proportion in the two experimental groups who quit watching at the beginning of the program, but rather the difference occurred primarily among those viewers who had seen more than 10 minutes, and who probably in most cases had seen the entire program. This suggests that, indeed, the design of the program had influenced the viewers' inclination to see the program in its entirety, and that the production "tricks" used to heighten motivation in the E-version were effective in this respect. Parallel with this main effect, an interaction effect occurred between

	L-sample (N = 112)	E-sample (N = 164)
Viewers of "Tooth by tooth", expressed as a proportion of the total TV audience that evening	72 %	70 %
The number who watched TV at all that evening (base figure)	(N = 81)	(N = 114)
The proportion watching		
more than 10 minutes	54 %	67 %
more than 5 minutes	80	83
more than 3 minutes	89	91
more than 1 minute	96	95

Table III:5A. The length of exposure. Effects study.

	L-sample		E-sample	
	N	%	N	%
Male	(51)	49	(57)	39
Female	(29)	41	(57)	28
Born 1930 -	(29)	45	(44)	32
Born - 1930	(51)	47	(70)	34
Elementary school	(37)	41	(68)	39
More than elementary school	(43)	51	(46)	26
Total	(80)	46	(114)	33

Table III:5B. Percentage of "Tooth by tooth" viewers who quit watching (saw less than 10 minutes). Effects study.

Test of significance for the difference between the versions in the different education groups:

		Elementary school			More than elem. school		
		L	E	Both	L	E	Both
Seen > 10 minutes	Yes	22	42	64	21	34	55
	No	15	27	42	22	12	34
		$Chi^2 \simeq 0$			$Chi^2 = 5.92$, $p < 0.025$		

the program format and the education level of the audience: the design of the program had no effect on the propensity of the Low-education groups to see the entire program, while among the Higher education groups, the E-version was the more attractive.

Thus, the E-version was not more attractive among those groups with less knowledge of the subject matter. On the contrary, the difference between the versions is most pronounced among those groups who generally are more selective in their television viewing habits, and who may be assumed to have had the least to learn from the program, i.e. those offered the least reward by the content. It should be noted that the differences observed in the length of exposure are most pronounced in the group who showed only a small difference in perception of the program versions in the Mail questionnaire survey. This relationship will be discussed in the closing portions of the report.

III:5b. The exposure activity variable

As mentioned earlier, in both the samples roughly three-fourths of the subjects (those viewing more than 5 minutes) viewed the program in the company of another person or persons. In both groups 65 % of these subjects discussed dental health or dental care in conjunction with exposure to the program. Thus, nothing indicates that the program format has had any effect as a stimulant of such activity as discussion of the content, or that such activity has to varying degrees influenced the effect of the two versions on other dependent variables.

The level of attention during exposure to the program is naturally difficult to ascertain by means of an interview, particularly if it is carried out after an interval of 1 - 2 days. The direct question as to how attentively the respondent felt he watched the program produced quite a skewed distribution, as had been anticipated. Only a few subjects reported having watched the program "less attentively",

12 % in the E-sample, and 11 % in the L-sample. The answers to this question were not related to whether or not one viewed the program alone or in company, but were clearly related to whether or not the subject engaged in some other activity during the broadcast.

The proportion who only watched television during the program was, as noted above, somewhat larger in the L-version. This difference may have some consequences for the comparison of the versions regarding effects on knowledge and attitudes. The more viewers who only watched TV (and in their own judgment watched more attentively), the greater the expected effect regarding learning and influences on attitudes.

III:6. RESULTS OF THE EFFECTS STUDY: KNOWLEDGE AND ATTITUDES

III:6a. The knowledge variable

The questionnaire included a number of statements about dental health and dental care, which the subjects were to characterize as correct or incorrect. In the instructions, it was noted that there was a correct answer to each of the statements. If the respondent did not know the answer, he/she was instructed to guess. A few subjects refused to respond to certain statements, but no statement distinguished itself in this respect. Tables III:6A and III:6B present the knowledge questions and the distribution of responses to them within various subgroups.

The original intention was to form an index on the basis of the responses to all the statements referring to program content. A check of the intercorrelations[1)] between all the items revealed a low average intercorrelation, however, and an estimated alpha-value (McKennell, 1970) of only 0.41. A closer examination revealed that particularly a few items (the four last items in Table III:6A) influenced the alpha value. One of these (No. 58) achieved an extremely skewed distribution of responses, while another (No. 71) dealt with an issue not treated in the program. A third item (No. 60) was expressed somewhat ambiguously - "disease of age" (ålderssjukdom) might be interpreted either as a disease generally afflicting elderly persons or as a disease arising because of changes involved in the aging process. When these four items were excluded, the alpha-value for the remaining items rose to 0.53. The analysis of the knowledge variable was performed using an index (the K-index) calculated on the basis of correct responses to these remaining questions.

---

1) Intercorrelations were calculated on the data from the two samples combined.

Item	Elementary school		More than elementary school	
	E-group (N = 55)	L-group (N = 30)	E-group (N = 39)	L-group (N = 35)
56. Practically all 7-year-olds have cavities in their teeth.	69	70	87	91
61. It may be pleasant to use a tooth-pick after a meal, but it doesn't help against tooth decay.	58	50	54	71
62. One should begin brushing children's teeth when they are about 6 years old.	80	90	97	89
64. One develops only half as many cavities if there is fluorine in the drinking water.	89	87	95	77
65. Inflammation of the gums is the final stage of paradontitis.	24	30	46	46
66. Loss of teeth is due to an infection which occurs because the teeth are not kept clean.	85	77	82	89
67. Most adults suffer from gum inflammation.	55	43	44	43
68. Paradontitis is a disease which cannot be diagnosed before the teeth begin to loosen.	62	43	79	77
69. The more often one eats during the day, the greater the risk of developing caries.	96	80	85	83
57. Persons in their twenties seldom have inflamed gums.	67	63	69	71
58. Tooth decay occurs because bacteria produce an acid which eats holes in the teeth.	96	97	95	97
60. Loss of teeth is primarily a disease of age.	76	87	85	71
71. If children's baby-teeth are not properly cared for, they will more easily develop caries in their permanent teeth.	20	23	28	23

Table III:6A. Responses to the knowledge questions in the respective Education groups. Percentage correct responses among those having seen at least 5 minutes of the program.

Item	Younger (born 1930 -)		Older (born - 1930)	
	E-group (N = 36)	L-group (N = 26)	E-group (N = 58)	L-group (N = 39)
56. Practically all 7-year-olds have cavities in their teeth.	89	96	69	72
61. It may be pleasant to use a tooth-pick after a meal, but it doesn't help against tooth decay.	67	73	50	54
62. One should begin brushing children's teeth when they are about 6 years old.	92	96	85	85
64. One develops only half as many cavities if there is fluorine in the drinking water.	92	76	92	85
65. Inflammation of the gums is the final stage of paradontitis.	36	58	31	25
66. Loss of teeth is due to an infection which occurs because the teeth are not kept clean.	92	96	80	75
67. Most adults suffer from gum inflammation.	50	53	50	39
68. Paradontitis is a disease which cannot be diagnosed before the teeth begin to loosen.	76	88	66	44
69. The more often one eats during the day, the greater the risk of developing caries.	94	88	90	77
57. Persons in their twenties seldom have inflamed gums.	86	81	62	59
58. Tooth decay occurs because bacteria produce an acid which eats holes in the teeth.	98	86	95	97
60. Loss of teeth is primarily a disease of age.	90	84	75	75
71. If children's baby-teeth are not properly cared for, they will more easily develop caries in their permanent teeth.	28	21	23	23

Table III:6B. Responses to the knowledge questions in different Age groups. Percentage correct responses among those having seen at least 5 minutes of the program.

III:6b. The attitude variables

Nine statements expressing attitudes were presented, to which the subjects could answer either "Agree entirely", "Agree somewhat" or "Disagree entirely". The attitude statements were presented prior to the knowledge questions, and the instructions pointed out that these questions concerned personal opinions. Tables III:6C and III:6D present the distribution of responses for these items.

In order to form an index, the various response alternatives were assigned the values 0, 1 or 2, the highest value being assigned to the alternative which corresponded to the attitude expressed in the program. The sum of the values for all the responses was divided by the number of statements responded to and multiplied by 100. Thus, the index varies between 0 and 200. The attitude index (A-index) also turned out to exhibit a low internal consistency (alpha ca. 0.40), which motivated a search for a subgroup of items with a higher average intercorrelation. Yet another factor called for this procedure in the case of the attitude statements, however. While the statements included corresponded to the content of the program, certain points in the program clearly received more emphasis or a more detailed treatment than did others. The over-all theme of the program was that caries as well as paradontitis are both diseases which the individual himself can combat in various ways. One might say that the program aimed to create a "preventive" attitude toward dental disease.

A closer examination of the intercorrelations revealed that four attitude statements reflecting this theme were highly interrelated. They were

Nos. 48. There is not much one can do about losing one's teeth in old age.

53. Everyone has cavities; there is not much one can do about it.

54. Pre-school children don't have to visit the dentist, since they get good dental care when they start school.

55. If you brush your teeth morning and night, you need not go to the dentist unless you feel something is wrong.

Statement	Younger (born 1930 -) E-group (N = 36)	Younger (born 1930 -) L-group (N = 26)	Older (born - 1930) E-group (N = 58)	Older (born - 1930) L-group (N = 39)
47. Personally, I feel that I know too little about how teeth should be cared for. (Agree entirely or Agree somewhat.)	72	77	70	77
48. There is not much one can do about losing one's teeth in old age. (Disagree entirely.)	72	53	56	43
49. I ought to go to the dentist more often than I do. (Agree entirely or Agree somewhat.)	49	53	57	59
50. If people only took proper care of their teeth, there need never be any cavities in them. (Agree entirely or Agree somewhat.)	80	65	76	68
51. Parents should be obliged to take their children to the dentist at 2 years of age. (Agree entirely or Agree somewhat.)	89	81	80	89
52. If you always brush your teeth carefully, you need not worry about losing them in old age. (Agree entirely or Agree somewhat.)	78	70	68	64
53. Everyone has cavities; there is not much one can do about it. (Disagree entirely.)	47	46	45	41
54. Pre-school children don't have to visit the dentist, since they get good dental care when they start school. (Disagree entirely.)	58	76	69	61
55. If you brush your teeth daily, morning and night, you needn't go to the dentist unless you feel something is wrong. (Disagree entirely.)	89	96	74	85

Table III:6C. Attitude statements among those who watched at least 5 minutes of the program. Percentage expressing the same attitude as was expressed in the program. (Responses advocated by the program within parentheses.)

Statement	Elementary school		More than elementary school	
	E-group (N = 55)	L-group (N = 30)	E-group (N = 39)	L-group (N = 35)
47. Personally, I feel that I know too little about how teeth should be cared for. (Agree entirely or Agree somewhat.)	72	84	69	71
48. There is not much one can do about losing one's teeth in old age. (Disagree entirely.)	62	37	64	57
49. I ought to go to the dentist more often than I do. (Agree entirely or Agree somewhat.)	56	67	49	43
50. If people only took proper care of their teeth, there need never be any cavities in them. (Agree entirely or Agree somewhat.)	81	74	72	60
51. Parents should be obliged to take their children to the dentist at 2 years of age. (Agree entirely or Agree somewhat.)	86	87	80	86
52. If you always brush your teeth carefully, you need not worry about losing them in old age. (Agree entirely or Agree somewhat.)	69	60	77	71
53. Everyone has cavities; there is not much one can do about it. (Disagree entirely.)	42	33	51	51
54. Pre-school children don't have to visit the dentist, since they get good dental care when they start school. (Disagree entirely.)	65	60	64	74
55. If you brush your teeth daily, morning and night, you needn't go to the dentist unless you feel something is wrong. (Disagree entirely.)	78	83	82	94

Table III:6D. Attitude statements among those who watched at least 5 minutes of the program. Percentage expressing the same attitude as was expressed in the program. (Responses advocated by the program within parentheses.)

A correlation matrix which included knowledge questions as well as attitude statements further showed that the above four items were related to several knowledge questions also having to do with the need for or value of preventive measures. These were

Nos. 62. One should begin brushing children's teeth when they are about six years old.

68. Paradontitis is a disease which cannot be diagnosed before the teeth begin to loosen.

60. Loss of teeth is primarily a disease of age.

These 7 items, which showed an internal consistency (alpha) of .58, were included in an index, designated PA (preventive attitude). The responses to the attitude statements were dichotomized, the alternatives corresponding to the attitude expressed in the program being assigned the value 1. Attitude statements Nos. 48 and 53, above, largely express a feeling of powerlessness in the face of the dental problems in question. They may be regarded as indicators of "external locus of control" in a specific context. In order to avoid comparisons of the experimental treatments in terms of individual items, these items were combined to an index designated EC (external control, alpha ca. .45).

The formation of indices on the basis of the data obtained is naturally somewhat questionable and, in the absence of cross-validation, the values for internal consistency indicated here cannot be taken as a measure of the reliability of the method of measurement. It is important, however, that the variables used in the analysis of results are so homogeneous as possible. The factual knowledge questions were selected entirely on the basis of the intercorrelation matrix, and the conceivable explanations given for the low correlation between the discarded items and the other items are obviously entirely ad hoc. As for the choice of attitude items, the point of departure has been their face validity with respect to the program content, but it should be pointed out that the four items in the PA-index form the most homogeneous cluster of all the attitude statements. The knowledge questions included in the PA-index have been chosen strictly on the basis of their

correlations with the attitude statements.

Finally, the EC-index consists of two items which, because of the nature of their content, should be given particular attention in the analysis of results. Since they correlate, it is appropriate for reasons of reliability to treat them together.

III:6c. Different comparisons

Length of exposure. The effect of the program on knowledge and attitudes must be considered to depend on how much of the program one has seen. As was discussed earlier, on the whole a larger proportion of the subjects quit watching the L-version, and this tendency was particularly marked among persons with higher education than elementary school. All comparisons of the two programs must thus be made between groups who have seen equally much of the program, holding the education variable constant. The implication of the comparisons varies, however, depending on where the time limit is set. If the limit is set at 10 minutes, the groups will presumably consist of subjects most of whom have probably seen the entire program, so that any differences between the groups in terms of knowledge, for example, may be attributed to the program design. The difference may depend on the fact that one version has had a greater impact on knowledge and/or that it was particularly attractive for persons with initially higher levels of knowledge. Given that those in the two samples who began watching the program have had the same distribution as regards the K-variable, the latter explanation may be discarded, if a good number of those who began watching have seen the entire program, and if this proportion is the same for both versions. If these conditions are satisfied, any differences in knowledge between viewers of the entire program in the two groups may be assumed to depend on the design of the program.

If, instead, one sets the time limit quite low and allows the experi-

mental groups to include all those who saw any part of the program, the relative ability of the two versions to hold an audience may also account for differences arising between the groups. It is reasonable to assume that the greater portion of its audience the program manages to keep, the greater the average knowledge of the group, provided the program has had any effect whatsoever. On the other hand, any difference between the versions must be dependent on to what extent those who quit watching the less popular version had particularly high or low knowledge. Thus, it is necessary to compare the versions with respect to those who quit watching as well.

The choice of the time limit must be made with reference to the purpose of the study, which, as has been discussed earlier, is to attempt to judge whether the design of the program has had any influence on the effect on knowledge and attitudes in a natural viewer situation. That situation is characterized by voluntary exposure, and the ability of the program to hold an audience has a decisive influence on its ability to affect knowledge and attitudes. The central question must thus be whether the groups who began to watch the respective program versions exhibit different average levels of knowledge, depending on which version of the program they have seen. Such a comparison must be made, however, taking into account both the distribution of the length of exposure of the respective groups and the knowledge and attitude of those portions of the samples who quit watching the program. In order to clarify the results further, however, the mean values of the groups who saw the entire program have been included.

Between-sample and within-sample comparisons. The versions can be compared in two principally different ways. First, one might compare the means of the two groups who saw something of the respective versions; secondly, one might compare the differences between the exposed and non-exposed groups in each of the two samples. While, as noted in Section III:4, the non-exposed groups in both samples had very similar means for the knowledge and attitude variables, certain differences were revealed. Those cases where a comparison between the mean values of the exposed groups shows a higher mean for one of the versions, while at the same time the difference between the exposed and the non-exposed groups is less for that version, quite

naturally present difficulties of interpretation. A fundamental condition for drawing any conclusions as to the differential effect of the versions will hereafter be that a version exhibits both a higher mean among the exposed groups and a greater difference between the exposed and non-exposed groups.

Table III:6E presents a summary of the differences between the portions of the samples who saw something of the program and the differences between the exposed and non-exposed groups for each of the program versions. The latter groups have been defined as those having television in the home who did not watch any television the evening of the experimental broadcast (see further Section III:7b, p. 136). In both of the education groups consistent results for three of the four dependent variables were obtained. The exceptions were the A-index among the low education groups and the PA-index among the higher educated. In both cases the difference between the versions is very slight. (It should be noted that a low value in the EC-index corresponds to the intended effect of the program.)

Tables III:6F and III:6G present summaries of the mean values for all the groups included in the further analysis. Excerpts from these tables are presented at different stages in the analysis of results. The two education categories are treated individually due to the interaction between the program version and level of education observed in connection with the length of exposure. As already mentioned, the primary comparisons are those made between the groups having seen <u>something</u> of the respective program version, but in order to further clarify the results, the results for those who watched more than 10 minutes of the program are also presented.

		XE - NXE	XL - NXL	XE - XL
Elementary school only				
K-index	(E)	.58	.20	.44
A-index		5	2	-1
PA-index	(E)	.49	-.09	.20
EC-index	(E)	.36	.07	.39
Higher education				
K-index	(E)	1.01	.75	.04
A-index	(L)	0	16	-1
PA-index		.52	.64	.04
EC-index	(L)	-.60	.25	-.03

Values based on Tables III:6F and III:6G.

X = All who saw something of the program (XE = E-version; XL = L-version)

NX = All who have a TV-set, but who did not watch television at all the evening of the experimental broadcast

(E) = The difference XE - XL is positive, and the difference X - NX is greater for E than for L.

(L) = The difference XE - XL is negative, and the difference X - NX is greater for L than for E.

Table III:6E. Summary of the differences between means of exposed and non-exposed groups, and between means for the two versions.

		Did not watch TV*	Seen < 10 min	Seen > 10 min	Seen something of program	Seen all three programs**
K-index	(E)	5.39	5.41	6.33	5.97	6.30
	(L)	5.33	5.46	5.57	5.53	6.00
A-index	(E)	122	129	129	127	120
	(L)	126	128	130	128	130
PA-index	(E)	4.32	4.59	4.95	4.81	4.99
	(L)	4.70	4.73	4.52	4.61	4.92
EC-index	(E)	.68	.46	.24	.32	.30
	(L)	.74	.40	.74	.61	.69

* Those who have television in their homes, but who did not watch TV the evening of the experimental broadcast.

** Program immediately prior, the experimental program, program immediately following.

Table III:6F. Summary of the mean values of various groups in the dependent variables. Subjects with elementary education only.

		Did not watch TV*	Seen < 10 min	Seen > 10 min	Seen something of program	Seen all three programs**
K-index	(E)	5.68	6.36	6.79	6.69	7.06
	(L)	5.90	6.32	7.00	6.65	6.70
A-index	(E)	127	125	128	127	120
	(L)	112	130	126	128	132
PA-index	(E)	5.12	4.91	5.88	5.64	5.38
	(L)	4.96	5.55	5.67	5.60	5.50
EC-index	(E)	.30	.55	.30	.36	.27
	(L)	.58	.36	.29	.33	.20

* Those who have television in their homes, but who did not watch TV the evening of the experimental broadcast.

** Program immediately prior, the experimental program, program immediately following.

Table III:6G. Summary of the mean values of various groups in the dependent variables. Subjects with higher education.

III:6d. Results in the Low-education group

The distributions of lengths of exposure among those with elementary school only are almost identical for the E- and L-samples, as shown in Table III:6H.

Those who began watching	E-group (N = 69)	L-group (N = 38)
Of these, the number who saw		
less than 5 minutes	20 %	21 %
5 - 10 minutes	19	18
more than 10 minutes	61	61

Table III:6H. Distribution of lengths of exposure among the subjects having an elementary education.

Thus, differences between the groups in terms of the dependent variables may not be attributed to differences in length of exposure to the program. Among those who quit watching, i.e. saw less than 10 minutes, the mean values for the dependent variables are practically identical (Table III:6J).

Subjects who watched < 10 min	E-group (N = 27)		L-group (N = 15)	
	M	s	M	s
K-index	5.41	1.99	5.46	1.63
A-index	129	32	128	37
PA-index	4.59	1.85	4.73	2.29
EC-index	.46	.63	.40	.71

Table III:6J. Means and standard deviations among those who watched less than 10 minutes.

Thus, neither is there reason to believe that self-selection, as expressed in length of exposure to the program, may have caused the groups to become initially non-comparable.

Analysis of the knowledge variable. The distribution of responses to the factual knowledge questions was presented in Tables III:6A and III:6B. Nine of these questions were included in the K-index (see p. 113). Table III:6K presents the mean values for the K-index.

As indicated earlier, the comparison of greatest interest is that between those who saw any part of the program at all, i.e. the top line in Table III:6K. It appears that those who saw the E-version show slightly higher means, but the difference is not significant even at the .10-level ($t = 1.60$, df 105, p ca. .12). The difference was somewhat greater among those who had watched at least 10 minutes, but it was not clearly significant ($t = 1.81$, df 63, p ca. .08). Thus, a certain tendency toward a greater effect of the E-version in the knowledge variable was revealed.

Analysis of the attitude variables. Tables III:6C and III:6D presented the distribution of responses to the attitude statements for various subgroups. The difference between the E- and L-groups did not form a clear, unambiguous pattern, but the differences were sometimes relatively large. This is true especially of items included in the PA-index, where the E-version consistently shows higher values.

As noted above, the A-index was formed by combining all of the statements. The mean values for this index are entirely identical for the two groups, regardless of where one sets the limits as to length of exposure. A closer analysis of this index is therefore unnecessary.

Table III:6L shows the results for the PA- and EC-indices. The E-groups shows a higher mean value in the PA-index, but the difference is not significant. For the EC-index (where a low mean value indicates a lower degree of external orientation), however, there is a significant difference in favor of the E-version both for those who watched any part of the program and for those who watched more than 10 minutes.

Persons with elementary education only who	E-group			L-group		
	M	s	N	M	s	N
saw something of the program	5.97	1.74	69	5.53	1.74	38
watched at least 10 min	6.33	1.46	42	5.57	1.81	23

Table III:6K. Means and standard deviations in two categories of the exposed subjects. K-index.

	E-group			L-group		
	M	s	N	M	s	N
PA-index:						
Seen something of the program	4.81	1.80	69	4.61	1.91	38
Watched > 10 min	4.95	1.75	42	4.52	1.61	23
EC-index:						
Seen something of the program	.32	.58	69	.61	.81	38
Watched > 10 min	.24	.53	42	.74	.85	23

Table III:6L. Means and standard deviations for the PA- and EC- indices. Subjects with elementary education only.

This result indicates that the E-version has to some extent had a greater effect on the general attitude which the program aimed to influence.

III:6e. Results in the High-education group

Among those with more than elementary school education, the lengths of exposure in the E- and L-groups were distributed as shown in Table III:6M.

As mentioned earlier, the E-version had a significantly greater ability to hold its audience among the more highly educated subjects. One would consequently expect that the E-group on the whole would show higher means, had the program had any effect.

Among those who quit watching the program, there is no difference between the versions for the K-index, while the L-version shows slightly higher values for the attitude variables (see Table III:6N). These differences are far from significant, however.

Analysis of the knowledge variable. The mean values for the K-index within the various categories of the High-education group are presented in Table III:6O. The differences between the versions are quite slight, and far from significant. Despite the fact that a much larger proportion of the E-group watched the entire program, this group has only a very slightly higher mean than the L-group when comparing all those who saw something of the program. As was shown in Table III:6E, however, a larger difference occurred between the exposed and the non-exposed subjects in the E-sample. The L-version shows a higher mean among those who watched more than 10 minutes, but the difference between these and the non-exposed subjects was of the same size for both the versions (see Table III:6G).

It is possible that the relatively few watching more than 10 minutes of the L-version were especially interested and consequently had higher initial knowledge. The effect of interest in the subject matter would be less pronounced if the two versions are compared for those who watched TV more or less continuously and saw something of both TV-News, the experimental program, and the subsequent program (cf. p. 137). As is shown in Table III:6G, the E-version has a higher mean when these subgroups of the two samples are compared.

Those who began watching	E-group (N = 45)	L-group (N = 43)
Of these, the proportion who saw		
less than 5 minutes	13 %	19 %
5 - 10 minutes	11	33
more than 10 minutes	76	48

Table III:6M. Distribution of the lengths of exposure among the subjects having higher education than elementary school.

Number who saw < 10 min	E-group (N = 11)		L-group (N = 22)	
	M	s	M	s
K-index	6.36	1.55	6.32	1.49
A-index	125	29	130	40
PA-index	4.91	1.50	5.55	1.64
EC-index	.55	.78	.36	.57

Table III:6N. Means and standard deviations among those who watched less than 10 minutes.

Subjects with more than elementary education who	E-group			L-group		
	M	s	N	M	s	N
saw something of the program	6.69	1.64	45	6.65	1.49	43
watched at least 10 min	6.79	1.66	34	7.00	1.41	21

Table III:6O. Means and standard deviations for two categories of the exposed subjects. K-index.

Analysis of the attitude variables. As was the case for the Low-education group, Table III:6E shows that the differences between the versions for the A-index and the PA- and EC-indices were negligible. The table also shows, however, that the differences between the exposed and non-exposed for all three indices were greater for the L-version. This will be examined in more detail in the next section of the report.

III:6f. Summary of the comparison of the versions

The results are on the whole consistent in that the version which has the highest mean among those who saw something of the program also shows a greater difference between the non-exposed and exposed groups. The differences between the versions are small, however. Only in the lower education group there is a clear tendency in favor of the E-version as regards knowledge and to some extent attitude, but even there the statistical basis for drawing conclusions is tenuous. Before commenting on the results a further analysis will be made in order to clarify the outcome.

III:7. THE ABSOLUTE EFFECT OF THE PROGRAMS

Regardless of whether any differences between the two versions in effects on knowledge and attitudes are revealed, some estimate of the extent to which the program may have influenced these variables should be made. If there are no differences between the two versions, it is clearly of interest to know if this is due to a universal lack of effect or to an equally strong effect. If there are differences between versions it is useful, at least from a practical point of view, to know if one of the versions had no effect at all. Especially when the analysis of differences between versions suffers from a good deal of uncertainty, it is necessary to find out whether a within-group analysis contradicts the results obtained. The question is thus: Were knowledge and attitudes of the viewers of the program affected to any significant degree?

At this point it is necessary to return for a moment to the purpose of the field study and determine which effect is relevant in this context, that is, the effect <u>of what</u>? In a normal laboratory situation exposure to the program in its entirety would probably be an independent variable, and the conclusion of the experiment would be something like, "If one presents this program to such-and-such a category of persons, the effect will be as follows ...". Such a question cannot be answered and is in fact not relevant in the present study, since in the natural mass media situation broadcasts are not "aimed" at any pre-selected category of persons. When the program began, there were a certain number of persons watching television. The question is, how these people on the average have changed with respect to their knowledge and/or attitudes as a result of the experimental broadcast. Thus, of interest

here is rather the effect of having been watching television and thereby having had the opportunity of seeing the program in the first place.

III:7a. The relationship between exposure and the dependent variables

The data in Figures III:7A and III:7B show a relationship in both the samples between having seen something of the program (exposure) and knowledge and attitudes (the K- and A-indices). It must now be determined whether such a relationship might have existed even if the program had had a totally different subject matter, i.e. whether or not those who watched the program had better initial knowledge about and more positive attitudes toward dental care and dental health. In the absence of a pre-measure no change can be observed. The only comparison possible, then, is a comparison of the exposed and non-exposed groups in the respective samples.

The comparison of the exposed with the non-exposed in this case should not be equated with the comparison normally made between experimental and control groups. There the groups are assumed to be interchangeable - any effect would be expected to be the same, had the experimental group acted as the control group, or vice versa. The groups and other experimental conditions are assumed to be the same in all respects except for exposure to the experimental treatment. In this case the situation is different - the only point in comparing the exposed groups with the non-exposed groups is to determine whether the former have changed because of the program. No conclusions whatsoever are drawn as to how the non-exposed groups might have been affected, had they seen the program. They can differ in many different respects from the exposed subjects, but given that both groups exhibit the same values for the dependent variable before the program was broadcast, data concerning the non-exposed can be used to judge whether or not the exposed subjects have been affected as a result of the broadcast.

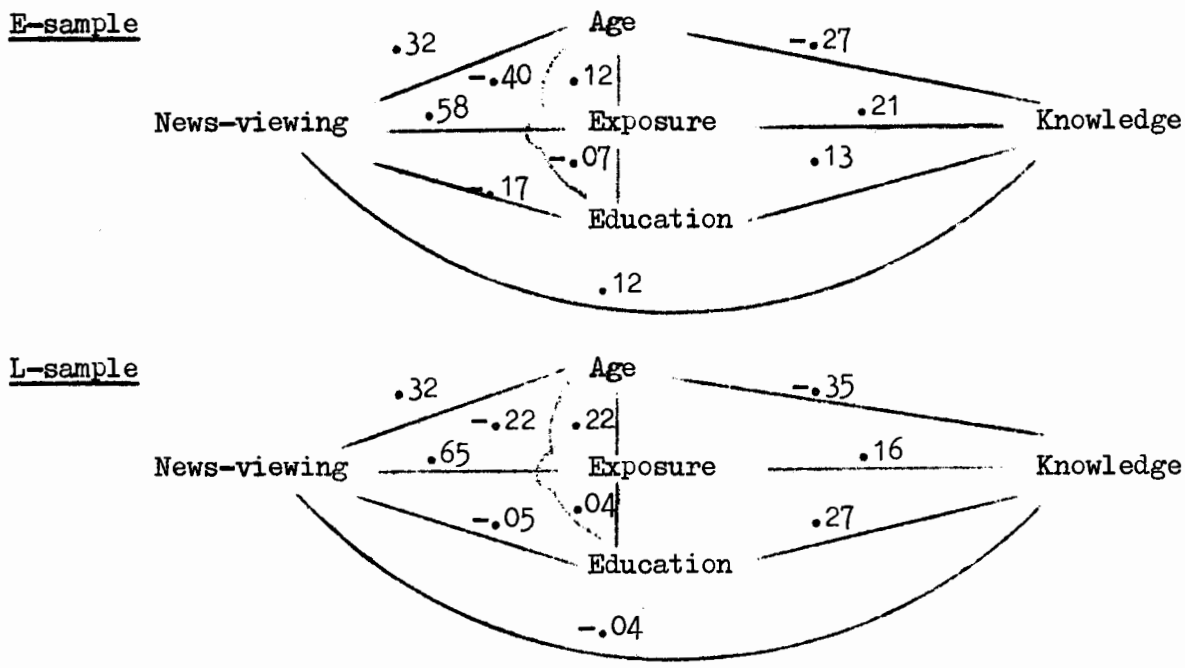


Figure III:7A. The knowledge variable. Correlations with exposure and control variables.

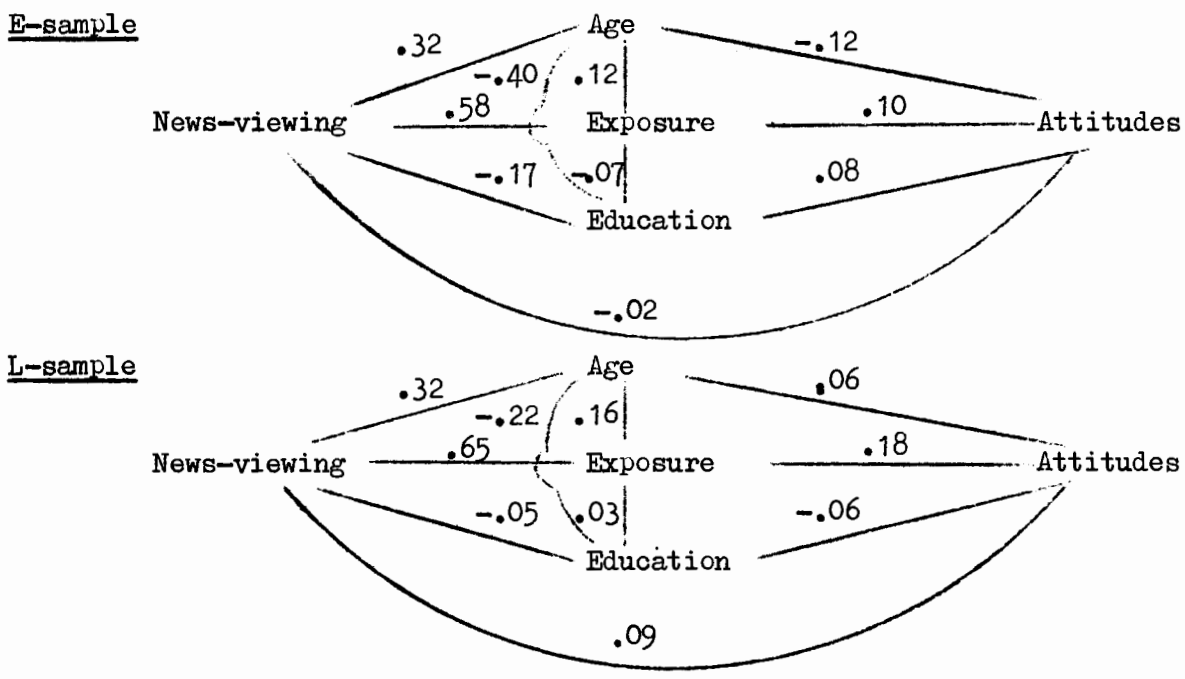
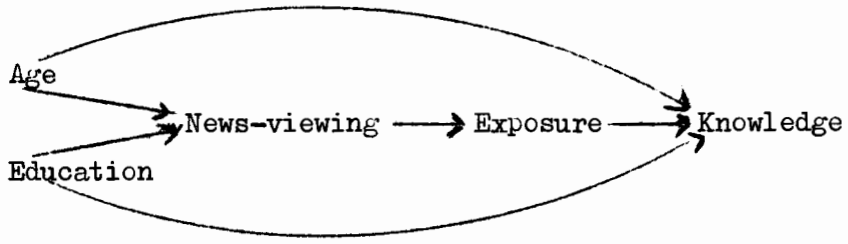


Figure III:7B. The attitude variable. Correlations with exposure and control variables.

This change need not depend on program characteristics alone, but may also arise from certain characteristics of the group exposed - the "confounding" of independent variables is intrinsic to the aim of carrying out the study in a naturally occurring situation.

The more detailed analysis will be confined to the knowledge variable. The first step must be to ascertain whether the relationship between knowledge and exposure can be considered spurious on the basis of the data at hand. If such is the case, there is no ground for further analysis.

As was discussed in Section III:5, age and education were related to exposure only indirectly, via the mediating variable News-viewing. Further, it turns out that the relationship between News-viewing and the K-index disappears when exposure is held constant. News-watching thus is not directly related to knowledge, but only indirectly. This is true of both samples, and the data obtained are in accordance with the following causal model:



The relationship between exposure and knowledge turns out to remain significant when age, education and news-viewing are held constant. The partial correlations are .15 for the E-sample and .32 for the L-sample. Thus, there is no reason to consider the relationship between exposure and knowledge spurious according to the above model.

The finding does not, of course, allow any conclusions as to the effect of exposure as such - the exposed groups may have had initially higher values in the knowledge variable. Before attempting to solve this problem, however, the importance of the observed relationship, that is, whether the exposure variable has any "explanatory value" should be examined. Should this turn out to be negligible, there is no point in performing a more detailed analysis.

One indicator of the explanatory value of the exposure variable is its contribution to the multiple determination coefficient when introduced last in a step-wise regression analysis (Darlington's, 1968, "usefulness"). The determination coefficients for the E- and L-samples respectively, with age, education, news-viewing and exposure as independent variables, are .1386 and .1995. The "usefulness" of the exposure variable is .0169 in the E-sample and .0289 in the L-sample.

These values can be tested for significance by means of an F-test (Darlington, 1968, p. 169), according to the following formula:

$$F = r^2_K\ (X,A,E,V) \qquad \left(\frac{N - n - 1}{1 - R^2}\right)$$

It turns out that the exposure variable yields a significant increment to the multiple determination coefficient in both samples: in the E-sample $F = 6.14$ ($p < .025$), in the L-sample $F = 7.26$ ($p < .01$). Thus it is desirable to judge to what degree this increment may be attributed to the effect of the programs, and whether there occurs any difference between the versions in this respect.

III:7b. <u>Comparisons between the exposed and non-exposed groups</u>

As pointed out earlier, the fundamental question is whether or not the exposed and non-exposed groups in the respective samples can be assumed to have differed initially in terms of the mean values in the dependent variables. Differences in other variables are not of interest here, as the aim is to attempt to judge the effect of the program under natural conditions, i.e. including self-selection and its correlates. The preceding analysis has given no indication that the exposed groups should have had initially higher mean values for prior knowledge than

the non-exposed. Rather the opposite is suggested by the fact that the background variable (age), which correlates most positively with exposure, correlates most negatively with knowledge.

A comparison of all non-exposed with all those who watched something of the experimental program shows, for the E-version, a significant difference in the means of the K-index for both educational groups ($p < .01$) and, for the L-version, a significant difference in the group with lower education ($p < .05$). Certain factors call for a more precise definition of the groups in making such a comparison, however. Those who were not exposed to the program consist of both those who have television in their homes, but who did not watch television at all the evening of the broadcast, and persons who saw one or more programs, but not the experimental broadcast. The more actively a person has chosen to see or not to see the program, the greater the likelihood that the selection is related to his prior knowledge of the subject matter. In order to reduce this effect, the non-exposed group has been defined as "those having television in their homes, but who did not watch TV the evening of the broadcast". A comparison of the non-exposed group (defined in the above manner) with the total sample of the Baseline study (see Appendix 1) shows no tendency toward lower mean values for the former group; the difference is in the other direction but is not significant (see Table III:7C). This difference is probably a context effect. In the Baseline study the questions were asked at the end of a long interview on TV-viewing during three days, while in the Effects study they formed part of a shorter interview and were included in a battery of similar questions. The difference may, however, have another cause. Immediately prior to the Baseline study a weekly magazine distributed to every household in the country a special issue which happened to contain an article on dental health. The effect of this information may have had greater impact at the time of the Effects study, which was carried out 2 - 6 days after the Baseline study.

	Non-exposed		Baseline study. Persons having TV in their homes
	E-sample	L-sample	
	(N = 158)	(N = 96)	(N = 304)
Persons with elementary education	3.12	3.00	2.78
Persons with higher education	3.32	3.35	3.14

Table III:7C. Means for knowledge questions included in the Baseline study.

Thus, the non-exposed groups do not differ from a nationally representative sample, interviewed 1 - 3 days prior to the broadcast of the experimental program.

The exposed groups can be compared in a similar manner. One might suspect, for example, that persons who saw only the experimental program may have had different prior knowledge of the subject matter than other viewers. Only a few belong to this category, as 80 % of those who saw any part of the program also saw "TV-News" immediately preceding, while 10 % watched the experimental program and the following program. The group which may be expected to show the least relationship between prior knowledge and exposure to the program are those who saw "TV-News", the experimental program and the following program. The exposed group in the following analysis consists of these "three-in-a-row viewers", who exhibited the lowest degree of selectivity.

Thus, the comparison is drawn between those who did not watch any television the evening of the broadcast and those who more or less continuously sat before the TV-set for roughly one hour, during which time the experimental program was broadcast. It does not seem probable that the exposed group, so defined, would have had a higher average value for the dependent variables. But the results must be considered uncertain due to the very few cases falling in this category.

The analysis is performed as above, holding education constant. It should be noted that this comparison between the exposed and non-exposed does not constitute an alternative to the earlier analysis of differences between the programs - even if only one of the versions shows a significant difference, this may not be interpreted to indicate that the version had a statistically greater effect than the other. On the other hand, it is naturally of practical importance if only one of the versions may be assumed to have had any effect whatsoever.

III:7c. Analysis of the knowledge variable

Table III:7D presents the means for the K-index in various subgroups.

It is apparent that the exposed groups have consistently higher values than the non-exposed, but the differences are significant only for the E-sample. It should be remembered that the number of cases is smaller for the L-version than for the E-version. Assuming that the observed difference is true, the power of the test is thus less for the L-version. At the same time the observed difference between the exposed and non-exposed, expressed in standard deviation units, is less for the L-version (cf. Cohen, 1969). Assuming that the L-version had had the same means and variances but that the number of observations were equal to that of the L-version, the difference between exposed and non-exposed would still not be significant, however ($t = 1.63$ for the lower educational group, $t = 1.62$ for the higher educational group). Thus, the results obtained cannot be accounted for by the difference in the number of cases for the two versions.

	E-sample			L-sample		
Elementary school only	M	s	N	M	s	N
Not watched TV	5.39	1.73	80	5.33	1.60	46
Watched all three programs	6.30	1.79	20	6.00	1.84	13
Difference	.91	(p = .05)		.67	(n.s.)	
	t = 2.02, df = 29*					
Higher education	M	s	N	M	s	N
Not watched TV	5.68	1.50	78	5.90	1.79	50
Watched all three programs	7.06	1.39	16	6.70	1.62	10
Difference	1.38	(p < .01)		.80	(n.s.)	
	t = 2.807, df = 23*					

* Due to the sizable difference in the number of cases in the groups compared, the number of degrees of freedom has been adjusted according to Hays (1963, p. 322). The error terms have been calculated without using pooled estimates. The test thus is conservative.

Table III:7D. Means and standard deviations for the K-index among exposed and non-exposed groups.

III:7d. Summary

Given the uncertainty associated with the analysis of the absolute effect of the programs, it may be said that the pattern of results observed agrees quite well with the analysis of the differences between the program versions. Only exposure to the E-version can be assumed to have had any absolute effect as regards the knowledge variable, and there is a clear tendency for this version to have a greater effect than the

L-version among those with elementary education only.

Among the higher education groups no differential effect of program format was revealed, but this comparison is somewhat dubious due to the unequal distribution of exposure times for the two versions. A comparison of the less selective groups who had watched three programs in a row shows a higher mean for the E-version, and there is also a significant difference for this version between exposed and non-exposed. Although no conclusion can be drawn as to a differential effect of the two versions in the higher education groups, the fact that the L-version appeared not to have any effect whatsoever is, of course, of practical interest.

A corresponding analysis of the differences between exposed and non-exposed (as they were defined above) was carried out for the three attitude variables. Means for the various groups are shown in Tables III:6F and III:6G. The only significant difference observed was for the EC-index in the lower education group of the E-sample. As was mentioned in Section III:6d this variable was also the only point where a difference between versions was observed.

III:8. DISCUSSION

III:8a. The pattern of results

The purpose of the study was to find out whether program form, as defined by audience perception, would have any effect on learning and attitude change in a normal mass media situation characterized by self-selected exposure and freedom to refrain from exposure in the course of the program. The two versions were intended to be equally informative, but to offer different kinds of rewards. The E-version was designed to be more immediately rewarding (Schramm, 1954), whereas the gratifications derived from the L-version would be more of a delayed nature (the instrumental value of the information presented).

On the whole, perception of the program versions showed the expected pattern, but the differences were only partly significant. (It should be noted, however, that the number of observations in the Mail questionnaire study is fairly small.) The interaction between program version and educational level as regards length of time of exposure seems to be in accord with the assumptions about the nature of gratifications offered by the versions. For the higher educational group, the program contained less new information and the tendency to watch the program in its entirety would be more affected by immediate gratifications derived from the exposure. This effect may be enhanced by a generally more selective pattern of viewing among the more highly educated groups. Thus, it is reasonable that the E-version kept a larger proportion of the initial audience than did the L-version.

In the group with lower education this difference between versions

would not be so readily expected, since the L-version assumedly offered relatively more gratification in terms of new information. Furthermore, this information may have had higher relevance to the viewers, since dental status is generally less satisfactory in this group (Johansson, 1971). Less selective TV-viewing habits may of course also be an important factor.

The distributions of exposure times do not quite agree with the results of the Mail questionnaire study, which showed smaller differences in program perception for the high than for the low educational group. A possible ad hoc explanation is that the viewers in the voluntary captive audience of the Mail questionnaire study felt that they had a role as representatives of the general public - they may have rated the program on the basis of expectations as to how others would perceive it, or how useful it would be to the "average person". It may be that a subject who himself is fairly well informed of the subject matter more easily falls into such a role, precisely because he does not perceive himself as a member of the relevant audience for the program. No matter how one may interpret these results, they do indicate that program ratings by a captive audience cannot be usefully employed to predict actual exposure behavior.

The differences between education groups as regards effects on knowledge are most probably related to initial level of knowledge of the subject matter, and to differences in information-seeking habits and experiences. In the lower education group the receivers received more new information, while they may be assumed to have less practice of information processing as compared with the more educated. Any effect of program form on the level of attention or involvement would therefore be expected to affect learning more in this group than among the more highly educated. As was seen in the analysis, the difference between versions is pronounced only in the low education group. The absolute effect of the program, however, seems to have been stronger in the higher education group - the difference between exposed and non-exposed is consistently larger in this group than among the less educated (cf. Tables III:7F and III:6G). This is of course in accordance with the hypothesis about differential information processing capacity.

The lack of difference between versions as regards learning in the high education group does not agree with the fact that only the E-version showed a significant difference in the within-version analysis comparing exposed and non-exposed. To some extent this may have to do with the different definitions of "exposed" - in the comparison between versions this group included all viewers who had seen something of the program, while in the comparison within versions "exposed" were defined as three-in-a-row viewers. If the comparison between versions is made for the less selective three-in-a-row viewers, the E-version does have a somewhat higher mean. The difference is not significant, however, and all that can be said is that for both versions the difference between exposed and non-exposed is larger among high than among low educational groups, and that there is no basis for concluding that the E-version was more effective than the L-version in the high education group as regards learning.

In the attitude variables none of the versions seems to have had an appreciable effect. Both within- and between-versions analyses, however, show differences in the EC-index in favor of the E-version among the less educated. Examination of the individual attitude statements also reveals that out of six statements which more or less clearly express a preventive attitude (Nos. 48, 50, 52 - 55), the E-version has a higher percentage of "preventive responses" in all except one (55). In the latter, there is a small difference in the opposite direction. No conclusion can be drawn from this, but there is a clear indication that, in the low education group, the E-version did affect somewhat the audience's beliefs about the usefulness or necessity of taking preventive action as regards dental health.

III:8b. Power

The fact that the effects which could be noted were only in some cases statistically significant must be related to the very few number of cases

in several parts of the analysis. The size of the total sample was determined by primarily economic considerations, but it was considered reasonable in view of an expected exposure frequency of 25 - 30 %. Due to the fact that the analysis had to be performed separately for the two education categories, and because the differences between the versions were smaller than expected, the power of the t-tests is low. In retrospect, it is clear that the sample should have been considerably larger, in view of the small differences observed.

In an experiment which aims to clarify the importance of theoretically well-defined variables it is naturally extremely important that even slight differences which may have consequences for the modification of the theory be detected. In the present study, however, the independent variable is neither conceptually nor operationally well-defined - it is a dichotomous, nominal level variable representing two alternative approaches to the practical creation of a television program. Each approach includes a number of details concerning program design which may be expected to influence the effect of the program.

The more details in a given program version tending either to increase or to decrease the intended effect of the program, the greater the differences to be expected between the versions. Only if a good number of these details in one of the versions influences the effect in one and the same direction, is there cause to attach any importance to the program design as defined in this general manner.

In other words, it is of little interest to ascertain statistically significant but very slight differences between the versions. The details characterizing a given version cannot be defined operationally other than in terms of the intended pattern of receiver perceptions. The versions used in the experiment might have been designed in many other ways and still have represented the two approaches which the experiment aimed to compare. Only if the differences in effect are substantial, they may be expected to obtain in another situation, where the "entertainment" and the "lecture" versions have another appearance.

If one regards a true effect which corresponds to one-half standard deviation as "substantial", there would have been a probability of approximately .80 to discover (at the .05-level) an effect of this size in the comparison between versions among the less educated who saw something of the program. (The estimation of power done on the basis of Cohen, 1969, p.17 ff.) For that comparison the power of the test may thus not be considered too low in view of the purpose of the study. For the high education group, however, the power of the test would have been rather low even with a substantial true effect, and in the within-version analysis the number of cases is small enough to make power low even with a relatively large true difference between means.

III:8c. Design

As has been mentioned earlier, the experimental design of this study is not entirely satisfactory, and the analysis of results must be based on assumptions which only partially can be tested. The most essential problem is the use of natural groups in such a way that group membership and the experimental treatment have been totally confounded. This, however, was unavoidable, as both practical and technical factors required that one of the versions be broadcast over a single area only. It would naturally have been more desirable, had each version been broadcast over several areas.

Some form of pretest aimed to ascertain the comparability of the two areas/samples in terms of the dependent variable would have eliminated some of the difficulties in interpreting the results. A pretest on a separate sample, however, would not have offered much more than the data available for the non-exposed groups and in the Baseline study. Rather, it would have been necessary to employ a panel technique in order to determine whether the groups were initially comparable. Apart from the methodological problems associated with panel studies,

it may be noted that this approach would have required larger resources than available for the present study. Even if the pretest might have been made somewhat less comprehensive than the posttest in terms of the length of the questionnaire, the posttest would have required a larger sample to permit an estimation of the relationships between the pretest, self-selection and program effects (cf. Belson, 1967; Haskins, 1968; Campbell & Stanley, 1963).

Given the limit of two areas of study, two circumstances emerge in retrospect as weaknesses with regard to the comparison of the two versions. The first is the absence of additional control variables, which might have formed the basis for a test of the initial comparability of the exposed groups, and especially continuous variables which might have been used as co-variates in an analysis of co-variance. The other is the use of random samples. A stratified sample based on a factorial design with, for example, sex, age and education as orthogonal factors would have allowed a more sensitive analysis by reducing the error variance.

The interpretation of the results would also have been facilitated had the absolute effect of the versions been more certain. One possibility would have been to draw a sample in each of the areas of investigation and to instruct the respondents to do something else during the time of the broadcast. They might be instructed to listen to the radio, or, in the present-day situation with two television channels, to watch the "other" channel. These samples would thus form control groups of persons "prevented" from being exposed to the experimental program, which could be compared with the samples who were <u>able</u> to see the program (cf. Towers et al, 1962).

III:8d. <u>Concluding remarks</u>

The present study has a clearly non-theoretical character in that it does not aim to test specified hypotheses derived from theory. It does,

however, attempt to apply certain theoretical generalizations in a natural situation and to test their importance in this context. The attempt to achieve "naturalness" is reflected both in the choice of methodology (field experiment) and in the design of the experimental material (guided by the producer's opinion as to what could be considered realistic alternative courses of action). The program design appears, at least in certain respects, to have had some influence on the effect of the program on exposure, knowledge and attitudes. The differences are relatively small, but they indicate that the measures taken in the production of the E-version to increase the viewer's attention and motivation did affect the effectiveness of the program.

In the course of the investigation the present author has at times been haunted by doubts whether a study of this nature is worth the effort. The amount of work involved in the planning, execution, and analysis may not be justified in view of the uncertainty of results and the weak theoretical grounds for interpretation. But disregarding this particular study and its shortcomings, the author feels fully convinced that without taking the trouble to do naturalistic experiments researchers on mass communication effects will not be able to develop ecologically more valid theories.

One reason is that it is necessary to avoid the "subject awareness" problem, but more important is the fact that effects of mass media content are inevitably a function of a complex combination of the information presented and the nature of gratifications derived from the exposure. A better understanding of mass communication effects, be it on learning and attitude change or other psychological variables, requires that these two factors are considered simultaneously - the so-called effects approach will have to be integrated with the uses-and-gratifications approach. To some extent this is possible in the laboratory although few attempts have been made to learn how dimensions of uses and gratifications can be realized and manipulated in a captive audience setting (cf. Lundberg, 1972). Self-selection and freedom of action on the part of the audience, however, as well as the habitual and often probably casual nature of mass media exposure would seem extremely difficult to include in a laboratory experiment. Furthermore,

the laboratory situation is in itself an "exposure situation" from which the subject derives certain gratifications which may completely override any attempt to realize, under controlled conditions, the characteristics of natural mass media exposure. This is not to say that laboratory studies would be without value in this context - on the contrary, they will always be the most efficient way to test specific hypotheses. But naturalistic experiments should play an important role as a complementary source of data which may affect the kind of laboratory research undertaken. Although data obtained in field studies may often be less neat and unequivocal than those obtained in the laboratory, the crucial point is that they are <u>different</u> data - they serve to bring to the theory new conceptualizations of the phenomena under study (cf. Section II:4). The value of empirical data in theory development is of course not confined to the testing of hypotheses - an equally important role of the data is to serve as food for thought. In the field of mass communication effects, there is undoubtedly a strong need for thoughtful speculation which can help to formulate hypotheses which are derived from basic characteristics of the mass communication process.

The points raised above mainly refer to the role of naturalistic experiments in introducing new variables and relationships at the receiver's end. Naturally, attempts to carry out experiments in a natural context will also force the investigator to pay attention to existing conditions at the sender's end. Awareness of values and goals guiding the professional communicators' behavior as well as more systematic knowledge of the actual media output form a relevant basis for identifying research problems and formulating hypotheses as regards effects. In fact, at the bottom of the present author's arguments lies the conviction that there is a need to envision the place of psychological research on effects within the total context of mass communication as a significant social phenomenon. Notions of the roles and functions of the mass media in society will have to guide theoretical and empirical work to a much higher degree than has so far been the case. Within such a frame of reference it is very probable that other effects than learning and attitude change as well as a wider variety of media functions at the individual level will be singled out for closer

examination (cf. Nowak, 1971c; Nowak et al, 1972; Fjaestad & Nowak, 1972). More interest in naturalistic experimentation in the psychological study of mass communication effects may stimulate and even force investigators to conceptualize their research problems in a broader perspective, thus promoting the development of not only ecologically valid but maybe also socially significant theories.

R E F E R E N C E S

Abelson, R.P., Aronson, E., McGuire, W.J., Newcomb, T.M., Rosenberg,M.J., & Tannenbaum, P.H. (Eds.) (1968). Theories of cognitive consistency: A source book. Rand McNally, Chicago.

Allen, V.L. (1966 a). Effect of knowledge of deception on conformity. Journal of Social Psychology, 69, 101-106.

Allen, V.L. (1966 b). Review: Attitude and attitude change. American Sociological Review, 31, 283-284.

Allport, G.W. (1935). Attitudes. In Murchison, C.A. (Ed.), A handbook of social psychology. Clark University Press, Worcester, Mass.

Argyris, C. (1968). Some unintended consequences of rigorous research. Psychological Bulletin, 70, 185-197.

Aronson, E., & Carlsmith, J.M. (1968). Experimentation in social psychology. In Lindzey, G., & Aronson, E. (Eds.). The handbook of social psychology, Vol. II. Addison-Wesley, Reading, Mass.

Aronson, E., Turner, J., & Carlsmith, J.M. (1963). Communicator credibility and communication discrepancy as determinants of opinion change. Journal of Abnormal & Social Psychology, 67, 31-36.

Bakan, D. (1969). On Method. Jossey-Bass, San Francisco.

Bakan, D. (1966). The test of significance in psychological research. Psychological Bulletin, 66, 423-437.

Bandura, A. (1965). Vicarious processes: a case of no-trial learning. In Berkowitz, L. (Ed.), Advances in experimental social psychology, Vol. 2. Academic Press, New York.

Barber, T.X., & Silver, M.J. (1968 a). Fact, fiction, and the experimenter bias effect. Psychological Bulletin, Monograph Supplement, Vol. 70, Part 2, 1-29.

Barber, T.X., & Silver, M.J. (1968 b). Pitfalls in data analysis and interpretation: A reply to Rosenthal. Psychological Bulletin, Monograph Supplement, Vol. 70, Part 2, 30-62.

Barker, R.G. (Ed.) (1963). The stream of behavior. Appleton-Century--Crofts, New York.

Bauer, R. (1964). The obstinate audience. American Psychologist, 19, 319-328.

Belson, W.A. (1967). The impact of television. Crosby Lockwood, London.

Berelson, B., Lazarsfeld, P.F., & McPhee, W. (1954). Voting. University of Chicago Press, Chicago.

Björkman, M. (1970). On the ecological relevance of psychological research. In Lindblom, P. (Ed.), Theory and methods in behavioural sciences. Scandinavian University Books, Stockholm.

Blalock, H.M. (1964). Causal inference in nonexperimental research. University North Carolina Press, Chapel Hill, N.Car.

Brehm, J., & Cohen, A.R. (1962). Explorations in cognitive dissonance. Wiley, New York.

Brock, T.C., & Becker, L.A. (1966). "Debriefing" and susceptibility to subsequent experimental manipulation. Journal of Experimental Social Psychology, 2, 314-323.

Brunswik, E. (1956). Perception and the representative design of psychological experiments. University California Press, Berkeley.

Camilleri, S.F. (1962). Theory, probability and induction in social research. American Sociological Review, 27, 170-178.

Campbell, D.T. (1969). Prospective: artifact and control. In Rosenthal, R., & Rosnow, R.L. (Eds.), Artifact in behavioral research. Academic Press, New York.

Campbell, D.T. (1963). Social attitudes and other acquired behavioral dispositions. In Koch, S. (Ed.), Psychology: A study of a science, Vol. VI. McGraw-Hill, New York.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. Psychological Bulletin, 54, 297-312.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In Gage, N.L. (Ed.), Handbook of research on teaching. Rand McNally, Chicago.

Canon, L. (1964). Self-confidence and selective exposure to information. In Festinger, L. (Ed.), Conflict, decision, and dissonance, Stanford University Press, Stanford.

Carlson, R. (1971). Where is the person in personality research? Psychological Bulletin, 75, 203-219.

Chapanis, A. (1963). Men, machines and models. In Marx, M.H. (Ed.), Theories in contemporary psychology. MacMillan, New York.

Chapanis, N.P., & Chapanis, A. (1964). Cognitive dissonance: 5 years later. Psychological Bulletin, 61, 1-22.

Cohen, A.R. (1964). Attitude change and social influence. Basic Books, New York.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. Academic Press, New York.

Cook, T.D., Bean, J.R., Calder, B.J., Frey, R., Krovetz, M.L., & Reisman, S.R. (1970). Demand characteristics and three conceptions of the frequently deceived subject. Journal of Personality and Social Psychology, 14, 185-194.

Cronbach, L.J., & Furby, L. (1970). How we should measure "change" - or should we? Psychological Bulletin, 74, 68-80.

Darlington, R.B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Deutscher, I. (1966). Words and deeds: social science and social policy. Social Problems, 13, 235-254.

Feather, N.T. (1967). A structural balance approach to the analysis of communication effects. In Berkowitz, L. (Ed.), Advances in experimental social psychology, Vol. 3. Academic Press, New York.

Fennessey, J. (1968). The general linear model: A new perspective on some familiar topics. American Journal of Sociology, 74, 1-27.

Festinger, L. (1964). Behavioral support for opinion change. Public Opinion Quarterly, 28, 404-417.

Fillenbaum, S. (1966). Prior deception and subsequent experimental performance: The "faithful" subject. Journal of Personality and Social Psychology, 4, 532-537.

Fishbein, M. (1966). The relationships between beliefs, attitudes, and behavior. In Feldman, S. (Ed.), Cognitive consistency. Academic Press, New York.

Fjaestad, B., & Nowak, K. (1972). Företagen och massmedia. Forum, Stockholm

Freedman, J.L. (1965 a). Confidence, utility, and selective exposure: A partial replication. Journal of Personality and Social Psychology, 2, 778-780.

Freedman, J.L. (1965 b). Long-term behavioral effects of cognitive dissonance. Journal of Experimental Social Psychology, 1, 145-155.

Freedman, J.L., & Sears, D.O. (1965 a). Selective exposure. In Berkowitz, L. (ed.), Advances in experimental social psychology, Vol. 2, New York.

Freedman, J.L., & Sears, D.O. (1965 b). Warning, distraction, and resistance to influence. Journal of Personality and Social Psychology, 1, 262-266.

Friedman, N. (1967). The social nature of psychological research. Basic Books, New York.

Fröjd, B., & Olsson, L. (1968). Variation av skrämsel och specificitet i ett TV-program - effekter på attityder, kunskaper och handling. Handelshögskolan i Stockholm (seminarieuppsats, specialkurs P) (Mimeo).

Gold, D. (1969). Statistical tests and substantive significance. The American Sociologist, 4, 42-46.

Golding, S.L., & Lichtenstein, E. (1970). Confession of awareness and prior knowledge of deception as a function of interview set and approval motivation. Journal of Personality and Social Psychology, 14, 213-223.

Greenberg, M.S. (1967). Role playing: An alternative to deception. Journal of Personality and Social Psychology, 2, 152-157.

Greenspoon, J., & Brownstein, A.J. (1967). Awareness in verbal conditioning. Journal of Experimental Research in Personality, 2, 295-308.

Greenwald, A.G., Brock, T., & Ostrom, T.M. (Eds.) (1968). Psychological foundations of attitudes. Academic Press, New York.

Hammond, K.R. (Ed.) (1966). The psychology of Egon Brunswik. Holt, Rinehart & Winston, New York.

Haskins, J.B. (1968). How to evaluate mass communications: The controlled field experiment. Advertising Research Foundation, New York.

Hays, W.L. (1963). Statistics for psychologists. Holt, Rinehart & Winston, New York.

Heise, D.R. (1969). Some methodological issues in semantic differential research. Psychological Bulletin, 72, 406-422.

Higbee, K.L. (1969). Fifteen years of fear arousal: Research on threat appeals: 1953-1968. Psychological Bulletin, 72, 426-444.

Holmes, D.S., & Appelbaum, A.S. (1970). Nature of prior experimental experience as a determinant of performance in a subsequent experiment. Journal of Personality and Social Psychology, 14, 195-202.

Holmes, D.S. (1967). Amount of experience in experiments as a determinant of performance in later experiments. Journal of Personality and Social Psychology, 7, 403-407.

Hood, T.C., & Back, K.W. (1971). Self-disclosure and the volunteer: a source of bias in laboratory experiments. Journal of Personality and Social Psychology, 17, 130-136.

Horowitz, I.A. (1969). Effects of volunteering, fear arousal, and number of communications on attitude change. Journal of Personality and Social Psychology, 11, 34-37.

Horowitz, I.A., & Rotschild, B.H. (1970). Conformity as a function of deception and role playing. Journal of Personality and Social Psychology, 14, 224-226.

Hovland, C.I. (1959). Reconciling conflicting results derived from experimental and survey studies of attitude change. American Psychologist, 14, 8-17.

Hovland, C.I. (Ed.) (1957). The order of presentation in persuasion. Yale University Press, New Haven, Conn.

Hovland, C.I., Janis, I.L., & Kelley, H.H. (1953). Communication and persuasion. Yale University Press, New Haven, Conn.

Hovland, C., Lumsdaine, A., & Sheffield, F. (1949). Experiments in mass communication. Princeton University Press, Princeton, N.J.

Hyman, H.H. (1954). Interviewing in social research. University of Chicago Press, Chicago.

Johansson, S. (1970). Den vuxna befolkningens hälsotillstånd. (Utkast till kap 8 i Låginkomstutredningens betänkande). Allmänna Förlaget, Stockholm.

Jourard, S.M. (1968). Disclosing man to himself. Van Nostrand, Princeton, N.J.

Kaplan, A. (1964). The conduct of inquiry. Chandler, Scranton, Penn.

Kelman, H.C. (1968). A time to speak: on human values and social research. Jossey-Bass., San Francisco, Cal.

Kelman, H.C. (1958). Compliance, identification and internalization: three processes of attitude change. Journal of Conflict Resolution, 2, 51-60.

Kelman, H.C. & Hovland, C.I. (1953). Reinstatement of the communicator in delayed measurement of opinion change. Journal of Abnormal and Social Psychology, 48, 327-335.

Kintz, B.L., Delprato, D.J., Mettee, D.R., Persons, C.E., & Schappe, R.H. (1965). The experimenter effect. Psychological Bulletin, 63, 223-232.

Kish, L. (1959). Some statistical problems in research design. American Sociological Review, 24, 328-338.

Klapper, J.T. (1960). Effects of mass communication. Free Press, Glencoe, Ill.

Lana, R.E. (1969). Pretest sensitization. In Rosenthal, R., & Rosnow, R.L. (Eds.), Artifact in behavioral research. Academic Press, New York.

Lana, R.E. (1964). The influence of the pretest on order effects in persuasive communications. Journal of Abnormal and Social Psychology, 69, 337-341.

Lana, R.E. (1961). Familiarity and the order of presentation of persuasive communications. Journal of Abnormal and Social Psychology, 62, 573-577.

Lazarsfeld, P.F., Berelson, B., & Gaudet, H. (1948). The people's choice. Columbia University Press, New York.

Leventhal, H., & Niles, P. (1964). A field experiment on fear arousal with data on the validity of questionnaire measures. Journal of Personality, 32, 459-479.

Levy, L.H. (1967). Awareness, learning, and the beneficient subject as expert witness. Journal of Personality and Social Psychology, 6, 365-370.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 68, 304-305.

Lord, F.M. (1963). Elementary models for measuring change. In Harris, C.W. (Ed.), Problems in measuring change. University of Wisconsin Press, Madison, Wis.

Luchins, A., & Luchins, E. (1965). Logical foundations of mathematics for behavioral scientists. Holt, Rinehart & Winston, New York.

Lumsdaine, A.A. (1963). Instruments and media of instruction. In Gage, N.L. (Ed.), Handbook of research on teaching. Rand McNally, Chicago.

Lundberg, D., & Hultén, O. (1968). Individen och massmedia. Norstedts, Stockholm.

Lundberg, D. (1972). The impact of mass medium credibility as a function of expected use of information. (In preparation). Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm.

Lykken, D.T. (1968). Statistical significance in psychological research. Psychological Bulletin, 70, 151-159.

McGinnis, R. (1958). Randomization and inference in sociological research. American Sociological Review, 23, 408-414.

McGuire, W.J. (1969). Suspiciousness of experimenter's intent. In Rosenthal, R., & Rosnow, R.L. (Eds.), Artifact in behavioral research. Academic Press, New York.

McGuire, W.J. (1968). The nature of attitudes and attitude change. In Lindzey, G., & Aronson, E. (Eds.), The handbook of social psychology. Vol. III, Addison-Wesley, Reading, Mass.

McGuire, W.J. (1966). Attitudes and opinions. Annual Review of Psychology, Vol. 17, Palo Alto, Calif.

McKennel, A. (1970). Attitude measurement: use of coefficient alpha with cluster or factor analysis. Sociology, 2, 227-245.

McNemar (1958). On growth measurement. Educational and Psychological Measurement, 18, 47-55.

McQuail, D. (1969). Towards a sociology of mass communication. Collier-MacMillan, London.

McQuigan, F.J. (1963). The experimenter: A neglected stimulus object. Psychological Bulletin, 61, 421-428.

Masling, J. (1966). Role-related behavior of the subject and psychologist and its effect upon psychological data. Nebraska symposium on motivation, 14, 67-103.

Meehl, P.E. (1967). Theory-testing in psychology and physics: a methodological approach. Philosophy of Science, 34, 103-115.

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. Human Relations, 18, 47-58.

Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67, 371-378.

Minor, M.W. (1970). Experimenter-expectancy effect as a function of evaluation apprehension. Journal of Personality and Social Psychology, 15, 326-332.

Morrison, D.E., & Henkel, R.E. (Eds.) (1970). The significance test controversy. Aldine, Chicago.

Morrison, D.E., & Henkel, R.E. (1969). Significance tests reconsidered. The American Sociologist, 4, 131-140.

Newcomb, T.M. (1943). Personality and social change. Holt, New York.

Nilsson, S. (1971). Publikens upplevelse av TV-program. Del I: En faktoranalytisk studie av programskattningar. Sveriges Radio, Avdelningen för Publik- och Programforskning. Rapport 175/70. (Mimeo).

Nilsson, S. (1970). Upplevelse av TV-program - försök med faktoranalys av programskattningar. Sveriges Radio, Avdelningen för Publik- och Programforskning. (Mimeo.)

Nordenstreng, K. (1969). Toward quantification of meaning. An evaluation of the semantic differential technique. Annales Academiae Scientarium Fennicae, Ser B. Tom. 161,2. Helsingfors.

Nowak, K. (1971 a). Effekter av meddelandets form. Ett fältexperiment rörande effekten av ett TV-programs utformning på kunskaper och attityder. Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm. (Mimeo).

Nowak, K. (1971 b). TV-producenters och publikens bedömningar av TV-program - en explorativ studie. Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm. (Mimeo).

Nowak, K. (1971 c). Från informationsteknik till kommunikationsteknik. In Mattsson, L.-G. (Ed.), Människor och företag i kommunikationssamhället. Prisma, Stockholm.

Nowak, K. (1968). Om teorier för köparpåverkan. Markedskommunikasjon, 5, 30-39.

Nowak, K. (1966). Två experiment rörande effekten av skrämselpropaganda. Studier i ekonomisk psykologi, Nr 28. Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm. (Mimeo).

Nowak, K., Carlman, B., & Wärneryd, K.-E. (1966). Masskommunikation och åsiktsförändringar. Norstedts, Stockholm.

Nowak, K., Julander, C.-R., Lundberg, D., & Ölander, F. (1972). Alternativa mål för samhällsinformation. Sociologisk Forskning (in press).

Nowak, K., & Wärneryd, K.-E. (1969). Kommunikation och påverkan. Prisma, Stockholm.

Orne, M.T. (1969). Demand characteristics and the concept of quasi-controls. In Rosenthal, R., & Rosnow, R.L. (Eds.), Artifact in behavioral research. Academic Press, New York.

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

Ostrom, T.M. (1968). The emergence of attitude theory: 1930-1950. In Greenwald, A.G., Brock, T., & Ostrom, T.M. (Eds.), Psychological foundations of attitudes. Academic Press, New York.

Riecken, H.W. (1962). A program for research on experiments in social psychology. In Washburne, N.F. (Ed.), Decisions, values, and groups. Vol. 2. Pergamon Press, New York.

Rozeboom, W.W. (1960). The fallacy of the null-hypothesis significance test. Psychological Bulletin, 57, 416-428.

Rosenberg, M.J. (1969). The conditions and consequences of evaluation apprehension. In Rosenthal, R., & Rosnow, R.L. (Eds.), Artifact in behavioral research. Academic Press, New York.

Rosenberg, M.J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 3, 113-123.

Rosenberg, M.J., & Abelson, R.P. (1960). An analysis of cognitive balancing. In Rosenberg, M.J., Hovland, C.I., McGuire, W.J., Abelson, R.P., & Brehm, J.W. (Eds.), Attitude organization and change. Yale University Press, New Haven, Conn.

Rosenthal, R. (1968). Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. Psychological Bulletin, Monograph Supplement, Vol. 70, Part 2, 30-47.

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

Rosenthal, R., Persinger, G.W., Vikan-Kline, L., & Mulry, R.C. (1963). The role of the research assistant in the mediation of experimenter bias. Journal of Personality, 31, 313-335.

Rosenzweig, S. (1933). The experimental situation as a psychological problem. Psychological Review, 40, 337-354.

Rosnow, R.L., & Rosenthal, R. (1970). Volunteer effects in behavioral research. In New Directions in Psychology, Vol. 4. Holt, Rinehart & Winston, New York.

Rosnow, R.L., & Rosenthal, R. (1966). Volunteer subjects and the results of opinion change studies. Psychological Reports, 19, 1183-1187.

Rosnow, R.L, Rosenthal, R., McConochie, R.M., & Arms, R.L. (1969). Volunteer effects on experimental outcomes. Educational and Psychological Measurement, 29, 825-846.

Rosnow, R.L., & Suls, J.M. (1970). Reactive effects of pretesting in attitude research. Journal of Personality and Social Psychology, 15, 338-343.

Rubin, Z., & Moore, J.C. (1971). Assessment of subjects' suspicions. Journal of Personality and Social Psychology, 17, 163-170.

Schramm, W. (1954). How communication works. In Schramm, W. (Ed.), The process and effects of mass communication. Urbana, Ill.

Schultz, D.P. (1969). The human subject in psychological research. Psychological Bulletin, 72, 214-228.

Scott, W.A. (1968). Attitude measurement. In Lindzey, G., & Aronson, E. (Eds.), The handbook of social psychology, Vol. II. Addison-Wesley, New York.

Selvin, H.C. (1957). A critique of tests of significance in survey research. American Sociological Review, 22, 519-527.

Sherif, C.W., & Sherif, M. (Eds.) (1967). Attitude, ego-involvement, and change. Wiley, New York.

Sherif, M., & Cantril, H. (1947). The psychology of ego-involvements. Wiley, New York.

Sherif, M., Sherif, C., & Nebergall, R. (1965). Attitude and attitude change. W.B. Saunders, Philadelphia, Penn.

Sigall, H., Aronson, E., & Van Hoose, T. (1970). The cooperative subject: Myth or reality? Journal of Experimental Social Psychology, 6, 1-10.

Silverman, I. (1968). Role-related behavior of subjects in laboratory studies of attitude change. Journal of Personality and Social Psychology, 8, 343-348.

Silverman, I., Shulman, A.D., & Wiesenthal, D.L. (1970). Effects of deceiving and debriefing psychological subjects on performance in data experiments. Journal of Personality and Social Psychology, 14, 203-212.

Sjöberg, L. (1971 a). Missnöjd med laboratoriet. Psykolognytt, 17 (5), 3-5.

Sjöberg, L. (1971 b). The new functionalism. Scandinavian Journal of Psychology, 12, 29-51.

Smith, N.C. (1970). Replication studies: A neglected aspect of psychological research. American Psychologist, 25, 970-975.

Snider, J.G., & Osgood, C.E. (Eds.) (1969). Semantic differential technique. A source book. Aldine Publ., Chicago.

Solomon, R.L. (1949). An extension of control group design. Psychological Bulletin, 46, 137-150.

Spielberger, C.D. (1965). Theoretical and epistemological issues in verbal conditioning. In Rosenberg, S. (Ed.), Directions in psycholinguistics. MacMillan, New York.

Stricker, L.J. (1967). The true deceiver. Psychological Bulletin, 68, 13-20.

Stricker, L.J., Messick, S., & Jackson, D.N. (1969). Evaluating deception in psychological research. Psychological Bulletin, 71, 343-351.

Stricker, L.J., Messick, S., & Jackson, D.N. (1967). Suspicion of deception: Implications for conformity research. Journal of Personality and Social Psychology, 5, 379-389.

Thurstone, L.L. (1931). The measurement of social attitudes. Journal of Abnormal and Social Psychology, 26, 249-269.

Towers, I.M., Goodman, L.A., & Zeisel, H. (1962). A method of measuring the effects of television through controlled field experiments. Studies in Public Communication, Vol. 4, 87-110. University of Chicago Press.

UNESCO (1970). Mass media in society. The need for research. Reports and Papers on Mass Communication, Nr 59. Paris.

Webb, E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L.B. (1966). Unobtrusive measures: nonreactive research in the social sciences. Rand McNally, Chicago.

Weick, K.E. (1969). Laboratory experimentation with organizations. In Cummings, L.L., & Scott, W.E. (Eds.), Readings in organizational behavior and human performance. Irwin & Dorsey, Homewood, Ill.

Weick, K.E. (1967). Promise and limitations of laboratory experiments in the development of attitude change theory. In Sherif, C.W., & Sherif, M. (Eds.), Attitude, ego-involvement and change. Wiley, New York.

Weick, K.E. (1966). Task acceptance dilemmas: A site for research on cognition. In Feldman, S. (Ed.), Cognitive consistency. Academic Press, New York.

Westling, S. (1967). Manual för BARBRO-systemet. Enskilda Utrednings-institutet, Taxinge.

Wicker, A. (1969). Attitudes vs. actions: The relationship of verbal and overt behavioral responses to attitude objects. Journal of Social Issues, 25, 41-78.

Winch, R.F., & Campbell, D.T. (1969). Proof? No. Evidence? Yes. The significance of tests of significance. The American Sociologist, 4, 140-143.

Wuebben, P.L. (1968). Experimental design, measurement, and human subjects: A neglected problem of control. Sociometry, 31, 89-101.

Wärneryd, K.-E. (1970). Can results from psychological experiments be generalized to situations outside the laboratory? In Lindblom, P. (Ed.), Theory and methods in behavioural sciences. Scandinavian University Books, Stockholm.

Zajonc, R.B. (1968). Cognitive theories in social psychology. In Lindzey, G., & Aronson, E. (Eds.), The handbook of social psychology, Vol. I. Addison-Wesley, Reading, Mass.

Zimbardo, P., & Ebbesen, E.B. (1969). Influencing attitudes and changing behavior. Addison-Wesley, Reading, Mass.

Ölander, F. (1969). Sambandet mellan attityder och beteende - ett mångfacetterat problem. Markedskommunikasjon, 6, 28-46.

A P P E N D I C E S

BASELINE STUDY

This study consisted of a number of questions attached to the daily interviews carried out by the Department for Audience and Programme Research (PUB) of the Swedish Broadcasting Corporation. The interviews were conducted by telephone on a sample of slightly less than 400 persons of the age 15 - 80; the total number of records was 371 (after weighting for those without telephones, one-third of whom were interviewed personally). The interviews were carried out during the period September 4 - 7, 1969, by the ordinary interviewer staff. The following items from the knowledge questions in the Effects study were included:

1. Persons in their 'twenties' seldom suffer from inflamed gums.

2. It may be pleasant to use a tooth-pick after a meal, but it doesn't help against tooth decay.

3. One should begin brushing children's teeth when they are about 6 years old.

4. Paradontitis is a disease which cannot be diagnosed before the teeth begin to loosen.

5. The more often you eat during the day, the greater the risk of developing caries.

The questions were asked at the end of the interviews, which otherwise dealt with what television programs the respondent had watched during the preceding three days. The following table shows the average number of correct responses in various subgroups (among persons who have TV):

Group	Number	Average number correct responses
Men	(147)	2.93
Women	(157)	2.92
Age		
- 24 years	(59)	2.97
25 - 44 years	(111)	3.08
45 - 64 years	(100)	2.91
65 - 80 years	(34)	2.38
Elementary education only	(181)	2.78
Some Secondary school	(94)	3.03
High-school graduates	(29)	3.48

INCIDENTAL TECHNIQUE SURVEY

This study was performed on two samples of approximately 100 persons, drawn from the census registers in Örebro and Linköping, respectively (age group 15 - 70). The interviews were conducted by telephone between the hours of 8.15 PM and 9.30 PM on September 8, 1969. (The program "Tooth by tooth" was broadcast 8.05 - 8.20.) The interviewers were students (in the fields of psychology, sociology or education) engaged by the Department for Audience and Programme Research of the Swedish Broadcasting Corp. and trained in interview technique. Persons without telephones were excluded from the samples. Approximately one week prior to the interview checks were made as to whether these persons actually had no telephones and whether the telephone numbers at hand were up-to-date.

The following table summarizes the outcome of the study as to the number interviewed:

	Örebro	Linköping	Total
No telephone	9	5	14
No answer (when telephoned)	12	13	25
Not at home (when telephoned)	15	11	26
Refused to participate	-	3	3
Changed place of residence (not in the population)	2	11	13
Excluded for other reasons	-	2	2
Total number not oontacted	38	45	83
Number of drop-outs in the sample	36	32	68
Sample within the population	101	97	198
Interviews completed	65	65	130

The content of the questionnaire used is indicated by the following summary of results. (The original questionnaire is included in the Swedish appendix to this report.)

Summary of results

	Örebro	Linköping	Together
Number who had the television set turned on during the evening of the broadcast	53	42	95
Number who saw any part of the experimental broadcast	34	30	64
Percentage of those who watched TV at all that evening	64 %	71 %	67 %
Number who saw less than 10 min of the experimental broadcast	16	14	30
Percentage of those who saw any part of the program	47 %	47 %	47 %
Number who left the room when they quit watching the program	13	9	22

Roughly one-third of those who saw nothing of the experimental broadcast had had the television set turned on at the time of the program. Among these, there was usually someone else in the household who watched the program (4 of 6 in the Örebro sample, 3 of 4 in Linköping).

Of those reporting watching the entire program, the vast majority report watching it "Quite closely" or "Very closely" (15/18 and 14/16 respectively). Among those who watched less than 10 minutes, roughly half report having watched the program "Less attentively" or "Inattentively" (6/10 and 6/14 respectively).

Those who watched the entire program discussed dental health or dental care more often with others (7/10 and 6/12 respectively) than did those who only saw a portion of the program (3/10 and 3/8 respectively).

Most of those who saw at least a portion of the program report having done nothing else during the program (25/31 and 19/30 respectively).

On the whole, these results agree with the corresponding results of the Effects study. The only difference of any greater consequence is that no correlation between time of exposure and program version appears in

the Incidental technique survey. The proportion who quit watching is the same for both versions, and due to the very small number of cases it is not possible to do an analysis within education groups (where the Effects study indicated differences).

MAIL QUESTIONNAIRE SURVEY

This survey involved two samples drawn from the census registers for the cities Örebro and Linköping. The samples comprised approximately 100 persons. The subjects received a letter from the Swedish Broadcasting Corp. asking them to take part in a study of how people perceive various television programs (see App. 3A:2). Roughly 15 % of the letters turned out to be undeliverable. One week later, Saturday, September 6, 1969, the rest of the sample was sent another letter containing instructions (App. 3B:1) and forms for rating the programs "Tooth by tooth" and "Gay time", to be broadcast the following Monday. The rating scales are listed in Section III:3b. These letters thus arrived in the sample households on Monday, the day of the broadcast.

As of Wednesday, September 10, 1969, completed and returned questionnaires numbered 40 from Örebro and 51 from Linköping. Several questionnaires were received later but were excluded, as they could have been filled out too long a time after exposure to the broadcast.

On Wednesday and Thursday (September 10 - 11) telephone interviews were conducted among those who had returned their completed questionnaires, 20 persons in Örebro and 25 in Linköping. Practically none of the subjects had found the program ratings difficult, and very few (4) indicated that they had talked with others about the program when doing the ratings.

Invitation letter for the Mail questionnaire study.

SWEDISH BROADCASTING CORPORATION

INVITATION

The Audience and Programme Research Department (PUB) of the Swedish Broadcasting Corporation conducts continuous studies in order to find out what television programs people watch and what they think of various programs. We are presently testing a new method, by which we ask a number of persons chosen at random to watch television during a certain time on a certain date. On the day assigned these persons will receive a questionnaire in the mail asking a number of questions as to what they thought of one or more of the programs broadcast during the specified time period.

The questionnaire takes only a few minutes to fill out, and the questions are very simple. Naturally, all responses will be handled completely confidentially and will be transferred to computer cards so that no individual's responses can be identified.

You have been included among those randomly chosen for the next survey. We would appreciate your help and participation by watching television between 8 PM and 9PM, Monday, September 8. You need only do two things: watch TV during the above hour, and fill in and return the questionnaire you will receive in the mail the day of the broadcast. If you do not have access to a television set, you need simply note this on the questionnaire and return it.

It is extremely important to us that as many of those in our sample as possible can participate in the study. We hope that you will be able to participate. You will hear from us again September 8.

With thanks in advance for your assistance,
Stockholm, September 1, 1969
Yours truly,
SWEDISH BROADCASTING CORPORATION
Dept. for Audience and Programme Research

Instruction letter for the Mail questionnaire study.

SWEDISH BROADCASTING CORPORATION

We hope you have received our previous letter. Enclosed you will find a questionnaire with questions concerning two TV programs which will be broadcast the same day you receive this letter, that is Monday, September 8. In order to participate in this survey you need only do two things:

1. Watch television this evening, Monday, September 8, between 8 and 9 PM.

2. Fill out and return the enclosed questionnaire with questions as to how you liked the programs. (The enclosed envelope is post-paid; you need not stamp it.)

Please read the instructions and the questions carefully before you watch the programs. As you can see, the questions concern your own personal and frank opinions of the TV-programs we are studying. Thus, you should answer according to your own feelings, without reference to what your family or other persons may think of the programs. There are no right or wrong answers. All that matters is your opinion. Please answer the questions as soon as possible after viewing the programs, and watch the programs just about as you normally watch TV.

If for some reason you cannot or do not wish to participate in the study, we would appreciate your writing your name on the blank questionnaire and returning it to us.

We hope you will participate, and we thank you in advance. If you wish any additional information, contact Mrs. Anne-Margrete Wachtmeister, tel.: 63 10 00, ext. 1134.

Yours truly,
Stockholm, September 5, 1969
SWEDISH BROADCASTING CORPORATION
Dept. for Audience and Programme Research

INSTRUCTIONS FOR THE QUESTIONS ABOUT THE TV-PROGRAMS

The purpose of this questionnaire is to get your opinion about a couple of television programs. The questionnaire is arranged so that you may express your opinion by marking various words which you think suitably describe the program.

The next page concerns the program "Tooth by tooth". There are a number of scales that look like this:

:____:____:____:____:____:____:____:

At either end of the scale appears a word which might be used to describe the TV-program, for example:

interesting :____:____:____:____:____:____:____: uninteresting

First decide which of the two words best fits in describing the program. Perhaps you think "interesting" fits better than "uninteresting". If you feel that "interesting" fits very well, then set an "x" in the space to the far left:

interesting :_x_:____:____:____:____:____:____: uninteresting

But, you may feel that the word "interesting" fits only quite well. In that case, you should place an "x" in the space next to farthest left:

interesting :____:_x_:____:____:____:____:____: uninteresting

If you feel that the word "interesting" fits only somewhat in describing the program, you should mark the third space from the left:

interesting :____:____:_x_:____:____:____:____: uninteresting

If, instead, you think that the word "uninteresting" suits the program better than "interesting", you should mark "x" in one of the three spaces to the right of center. If "uninteresting" fits very well, then mark to

the far right; if it fits quite well, then next to the space farthest right; if it fits somewhat, then mark the third space from the right.

FIRST CHOOSE WHICH OF THE TWO WORDS YOU FEEL FITS BEST.
THEN, ACCORDING TO YOUR OWN OPINION, MARK ON THE SCALE HOW WELL THIS WORD FITS IN DESCRIBING THE PROGRAM.

If you should decide that neither of the two words appropriately describes the program, mark "x" in the space in the center.

Indicate your opinion by marking each of the scales in this way. Only your personal opinion is of interest here. Don't pay any attention to what your family or other persons may think of the program. There are no right or wrong answers. All that matters is your personal opinion.

When you have completed the questions concerning "Tooth by tooth", continue in the same way for the questions about "Gay time".

PRODUCER STUDY

A circular letter was distributed through the internal mail at the Swedish Broadcasting Corp., inviting 35 television producers to participate in a study. The producers were primarily involved in the production of informative programs. The letter described the field experiment and the general scheme of the Producer study. Four separate opportunities were provided for the producers to view the programs, their choice as to which of the showings being left free.

Both program versions were shown at each of the showings, but the order of presentation was varied. The distribution of the audience between the various showings was quite uneven, however - for practical reasons it was impossible to assign a definite showing to each of the producers - so that 12 persons saw one order of presentation while only 6 saw the other. The subjects were distributed according to age and length of employment as active producer, as follows:

	Age			
	- 25 years	26 - 35 years	36 - 45 years	45 + years
Order of presentation I	1	7	3	1
Order of presentation II	3	2	1	-

	Years as active producer			
	- ½ year	½ - 1 year	2 - 5 years	5 + years
Order of presentation I	3	4	2	3
Order of presentation II	1	2	2	1

Prior to the showing of the first version, a detailed explanation of the types of judgments the producers were expected to make was given. The rating scales were exactly the same as in the Mail questionnaire study. First, they were told to make judgments of the program as they personally perceived it, followed by a judgment as to how they anticipated the public

would react to it on the average. They were told at which time of the evening the program had been broadcast. The rating scales were distributed immediately following the showing of the first version, and the producers filled them out. When this had been completed, the second version was shown, and the corresponding ratings were made.

The participants were not informed as to the results of the field experiment or as to the aims underlying the design of the respective program versions. These questions were discussed, however, following the showings and ratings.

Table III:3G in the report presents the ratings given the programs by the producers. See also Nowak (1971b, included in the Swedish appendix to this report). The following is an extract from the English summary of Nowak (1971b):

> "18 professional TV producers watched both versions of the program and made their judgments on the same set of scales, first indicating their own perceptions of the program, second indicating their expectations as to audience perceptions. Three analyses were made: How do the producers believe their own judgments differ from those of the audience, how do they actually differ, and how do producers' expectations as to audience perceptions differ from the actual audience perceptions?
>
> Since the producers had watched both versions of the program, whereas the audience groups had seen only one, the absolute differences in judgments are not comparable. Instead, the purpose was to analyze to what extent the differences between producers and audience <u>vary</u> as a function of program version and judgmental dimension.
>
> Results show that expected as well as actual differences between producers and audience vary significantly between the various judgmental dimensions (main effect of Dimension), and that they vary between the versions depending on which judgmental dimension is considered (interaction effect Dimension x Version). Expected and actual differences do not show the same pattern, however, but show different direction for some dimensions. Producers did not differ from the audience the way they expected.
>
> The producers' ability to predict audience perceptions is not different for the two versions or for different age or education groups in the audience samples, but again there is a significant effect of judgmental dimension and interaction Dimension x Version. In some dimensions they underestimate, in other they overestimate audience ratings. Especially they tend to underestimate audience appreciation (Evaluation and Attraction) of the lecture-oriented version.

In spite of these differences between producers and the audience as to judgments of each of the two versions, results show that both groups show the same pattern of perceived differences between the versions. In other words, even if the ability of producers to predict audience perceptions of a single version is not very high, they may very well be capable of choosing the version which best corresponds to their intentions as to audience perceptions."

EFFECTS STUDY

This study was conducted during the period September 9 - 11, 1969, on two samples: (1) a nationally representative sample of 400 persons, drawn by the Central Bureau of Statistics at the request of the Swedish Broadcasting Corp., and (2) a sample of 300 persons in the city of Örebro, drawn by systematic sampling of the census register of Örebro. In both cases the samples consist of a population of persons between the ages of 15 - 80 years.

In the national sample the interviews were conducted by members of the staff of interviewers employed by the Department for Audience and Programme Research (PUB), stationed in 13 different districts. The interviews in Örebro were carried out by interviewers who, although not a part of PUB's regular staff, had previously conducted surveys for PUB. They were students in behavioral science, with training in interview technique. Each of these interviewers received extensive instructions from the project leader; the regular staff in the districts were instructed during a training course in Stockholm.

The interviews were conducted by telephone. Respondents who did not have a telephone were sent telegrams asking them to phone the Swedish Broadcasting Corp., for which they would be reimbursed. Sixteen such interviews occurred in the national sample, and eight in the Örebro sample.

After exclusion of those who turned out not to belong to the population (non-Swedish speaking, deaf, deceased, etc.) the sample consisted of 393 persons in the national sample, and 246 in the Örebro sample. The large number of exclusions in the Örebro sample was due to the fact that the sample was drawn on the basis of 2-year-old data, and all those who had changed their address during this interval must be excluded (unless they moved locally within Örebro), as they no longer belonged to the area of the experiment. The number of drop-outs in the national sample amounted to 9 %, and in the Örebro sample it was 10 %.

The questionnaire forms were checked and coded under the supervision

of the project leader. Data processing was preceded by checks to ensure that no logical errors existed in the collected data. The questionnaire is included in the Swedish appendix to this report.

STUDY OF INFORMATION-SEEKING BEHAVIOR

A number of persons included in the sample for the Swedish Broadcasting Corp.'s interviews about TV viewing during the day of the experimental broadcast were sent an offer of information material from a national association promoting dental health (Tandvärnet). The materials were offered free of charge. The aim of the study was to determine whether exposure to the program "Tooth by tooth" was at all related to a willingness to order such information material. As is indicated in the summary below, a total of 274 offers were sent out (only to persons born 1914 or thereafter). The mailing occurred approximately one week after the TV-program, around September 15. Orders requesting material were received for more than one month thereafter. It is reasonable to assume, however, that any effect the program may have had would be reflected best among those who ordered the material immediately or shortly after the offer was made. Thus the Table distinguishes between the orders received prior to October 1 and those received later.

As appears in the Table below, the proportion of those ordering material prior to October 1 was greater among those who had watched the program "Tooth by tooth" than among those not exposed to the program. One might also compare the frequency of exposure among all those who received the offer with that among those who actually ordered material. The latter frequency turns out to be considerably higher, particularly if one includes only those orders received prior to October 1.

Results of the Information-seeking study

Offers sent out	274
Of these, the number who had watched "Tooth by tooth"	82
Total number of orders received	55
From persons who had seen "Tooth by tooth"	20
From persons who did not see "Tooth by tooth"	29
From persons for whom data is lacking	6

	N	Number who ordered material: Before 1/10	Number who ordered material: 1/10 or later	Total
Exposed	82	19 (23 %)	1 (1 %)	20 (24 %)
Non-exposed	192	21 (11 %)	8 (4 %)	29 (15 %)

Frequency of exposure among those who received the offer	30 %
Frequency of exposure among those ordering material (20/49)	41 %
Frequency of exposure among those ordering material prior to October 1 (19/40)	48 %

The relationship between exposure to the program and ordering material is significant at the .10-level ($p \approx .08$, $Chi^2 = 2.945$) among those who ordered the material prior to October 1. This may, of course, not be interpreted to indicate any effect of the program without a closer analysis of the various groups. No data for such an analysis are available. Interpretation would be rather difficult, anyway, considering the scant number of orders received.

The analysis carried out in Section III:7 of the report, however, does give some plausibility to the hypothesis that exposure to the program affected the willingness to order information materials.

STUDIER I EKONOMISK PSYKOLOGI

1. Wärneryd, K.-E. Ekonomisk psykologi. Stockholm: Natur och Kultur, 1959. 220 s. Ny reviderad och utvidgad upplaga, 1967. 415 s.

2. Wärneryd, K.-E. Undersökning av informativ och suggestiv annonsering. EFI, 1958. Stencil, 30 s.

3. Wärneryd, K.-E. (med biträde av Carlsson, R. och Ölander, F.) Alkoholkonsumtionens inriktning. EFI, 1958. Stencil, 177 s. Utg.

4. Carlsson, R. Synpunkter på en indirekt metod för bestämning av en produkts sociala anseende. EFI, 1960. Stencil, 91 s. (jämte bilagor).

5. Ölander, F. Mätning av preferenser med kvotskattning och parvisa jämförelser. EFI, 1961. Stencil, 51 s.

6. Wärneryd, K.-E. Ungdomens alkoholvanor. En konsumtionspsykologisk studie. EFI, 1961. Stencil, 237 s.

7. Ölander, F. Fraskomplettering som attitydmätningsinstrument. EFI, 1961. Stencil, 54 s.

8. Wärneryd, K.-E. Bilägaren och bilköpet. En modell och några intervjuresultat. EFI, 1961. Stencil, 124 s.

9. Wärneryd, K.-E., Carlsson, R. & Ölander, F. Ett begreppsschema för reklamforskning. Det Danske Marked, 1962, 21, s. 217-227. Särtryck kan erhållas från EFI.

10. Wärneryd, K.-E., Carlsson, R. & Ölander, F. Psykologisk forskning inom reklamen. En inventering av resultat och forskningsuppslag i anslutning till ett begreppsschema för reklamforskning. Det Danske Marked, 1962, 21, s. 365-380. Särtryck kan erhållas från EFI.

11. Wärneryd, K.-E. Studiet av konsumentbeteende inom de ekonomiska och psykologiska vetenskaperna. EFI, 1962. Stencil, 23 s.

12. Wärneryd, K.-E. Opinionsledare i samhället. EFI, 1962. Stencil, 18 s.

13. Ölander, F. Preferensteori och preferensmätning. Diskussion kring en föreställningsram och ett försök att påverka preferenser. EFI, 1962. Stencil, 109 s.

14. Wärneryd, K.-E. Påverkan av konsumentbeteende. Ingår i Boalt, C. & Jonsson, E. Konsumtionen i sociologisk belysning, s. 57-77. Stockholm: Svenska Bokförlaget/Norstedts, 1964.

15. Wärneryd, K.-E. The use of scaling and standardized indirect methods in consumer interviewing. Ingår i Industrial Business Psychology, Proceedings of the XIV International Congress of Applied Psychology, vol 5, s. 161-171. Köpenhamn: Munksgaard, 1962.

16. Wärneryd, K.-E. Amerikansk masskommunikationsforskning. Några intryck från en studieresa. EFI, 1963. Stencil, 35 s.

17. Nowak, K. Masskommunikationsforskning i Sverige. En översikt och bibliografi. Stockholm: Norstedt & Söner, 1963. 98 s.

18. Carlsson, R. Produkters sociala anseende. En skiss till teoretisk bakgrund och till mätmodell. EFI, 1964. Stencil, 149 s.

19. Ölander, F. Engelsk forskning i ekonomisk psykologi. Intryck från en studieresa samt en litteraturöversikt. EFI, 1964. Stencil, 82 s.

20. Nowak, K., Sandell, R.G. & Wärneryd, K.-E. Företagets externa kommunikation. Några artiklar kring ett tema. EFI, 1964. Stencil, 70 s.

21. Sandell, R.G. Konsumentbeteende ur teoretisk synvinkel - en diskussion och orientering. EFI, 1965. Stencil, 60 s.

22. Thorslund, S. Metoder vid för- och eftermätning av kommunikationseffekter. EFI, 1966. Stencil, 43 s.

23. Sandell, R.G. Likheten mellan väljare och valobjekt som allmänt valkriterium. EFI, 1966. Stencil, 81 s.

24. Remstrand, L.-G. Informationsspridning och opinionsbildning i trafiksäkerhetsfrågor. EFI, 1966. Stencil, 170 s.

25. Carlman, B. Perception av meddelanden. Några experimentella studier. Se s. 230-258 i Studier i ekonomisk psykologi: 32.

26. Carlman, B. Kategorisering av meddelanden. Tre experiment. Se s. 258--277 i Studier i ekonomisk psykologi: 32.

27. Nowak, K. Direkt och indirekt generalisering av attitydförändringar. EFI, 1967. Stencil, 46 s.

28. Nowak, K. Två experiment rörande effekten av skrämselpropaganda. EFI, 1967. Stencil, 87 s. Utg.

29. Seipel, C.-M. Kommunikatoreffekter och meddelandets argumentering. EFI, 1967. Stencil, 244 s.

30. Bjuvman, A., Schött, M., Björkman, J. & Herbst, K. Perception av reklamstimuli. EFI, 1966. Stencil, 86 s.

31. Stålberg, L., Säve-Söderbergh, B., Lindhoff, H. & Nauckhoff, F. Bedömning av produkter och personer. Två tillämpningar av kongruitetsprincipen. EFI, 1965/66. Stencil, 101 s.

32. Nowak, K., Carlman, B. & Wärneryd, K.-E. Masskommunikation och åsiktsförändringar. Stockholm: Norstedt & Söner, 1966. 371 s.

33. Ölander, F. & Hultén, O. Förteckning över datamaskinprogram för tillämpning inom pedagogik-psykologi-sociologi. EFI, 1967. Stencil, 34 s.

34. Ölander, F. & Seipel, C.-M. Sparbeteende ur psykologisk synvinkel. EFI, 1967. Stencil, 124 s.

35. Wärneryd, K.-E. & Nowak, K. i samarbete med Carlman, B. & Lindhoff, H. Mass communication and advertising. Stockholm: EFI och AB Svenska Telegrambyrån, 1967. 119 s.

36. Lundberg, D. & Hultén, O. Individen och massmedia. Stockholm: Norstedt & Söner, 1968. 426 s.

37. Nowak, K. & Wärneryd, K.-E. Kommunikation och påverkan. Stockholm: Prisma, 1969. 151 s.

38. Seipel, C.-M. Konkurrens med presenter och tjänster. En studie av konsumenters reaktioner på konkurrensåtgärder av gåvokaraktär. EFI, 1970. Stencil, 126 s.

39. Seminarieuppsatser i ekonomisk psykologi 1965-1970. Sammanställning och sammanfattningar av uppsatser som framlagts vid Handelshögskolan i Stockholm inom specialkursen ekonomisk psykologi åren 1965-1970. EFI, 1970. Stencil, 91 s.

40. Ölander, F., Hjelmström, E., Lillieskiöld, J. & Persson, A. Konsumenters beteende i smakprövningar av blindtest-typ jämfört med verbalt uttalade märkespreferenser. EFI, 1970. Stencil, 33 s.

41. Seipel, C.-M. Premiums - forgotten by theory. EFI, 1970. Stencil, 22 s.

42. Nowak, K. Massmedia och människorna i morgondagens samhälle. Ingår i Marknad och Media, s. 105-122. Stockholm: Norstedt & Söner, 1967. Särtryck kan erhållas från EFI.

43. Wärneryd, K.-E. Masskommunikationsforskning och mediaanalys. Ingår i Marknad och Media, s. 85-103. Stockholm: Norstedt & Söner, 1967. Särtryck kan erhållas från EFI.

44. Wärneryd, K.-E. Utnyttjande av massmedia i trafiksäkerhetsarbetet. Ingår i Andréasson, R., Halldin, M. & Lindgren, S. (red.): Människan i trafiken, s. 232-262. Stockholm: Natur och Kultur, 1967.

45. Ölander, F. Ekonomisk psykologi - ett nytt forskningsfält för psykologer. Nordisk Psykologi, 1967, 19, s. 197-202. Särtryck kan erhållas från EFI.

46. Nowak, K. Mass communication research in Sweden. Stockholm: Sveriges Radio, Audience Research Department, 1968. 14 s.

47. Lundberg, D. Svensk massmediaforskning - med speciell hänsyn till pressmedia. Stockholm: TU:s Förlag, 1969. Stencil, 31 s.

48. Ölander, F. Två artiklar om sparande och sparbeteende. Svensk Sparbankstidskrift, 1968, 52, s. 145-150 och 241-250. Särtryck kan erhållas från EFI.

49. Ölander, F. Studier av hushållens sparande - bör psykologerna medverka? Ingår i Lundberg, E. & Backelin, T. (red.): Ekonomisk politik i förvandling, s. 99-112. Stockholm: Norstedt & Söner, 1970.

50. Wärneryd, K.-E. Socialpsykologisk forskning. Ingår i 20 års samhällsforskning, s. 214-226. Stockholm: Norstedt & Söner, 1969.

51. Wärneryd, K.-E. Konsumentupplysningens verkningar. Ingår i Lundvall, L. (red.): Konsumenten och samhället, s. 33-43. Stockholm: Rabén & Sjögren, 1970.

52. Hultén, O. Nordisk dokumentationscentral för masskommunikationsforskning. Stockholm: Sveriges Radio, 1969. Stencil, 22 s. Finns även i engelsk version.

53. Fjaestad, B. & Jeleby, H. Dimensioner vid nyhetsvärdering. Stockholm: TU:s Förlag, 1970. Stencil, 67 s.

54. Ölander, F. & Seipel, C.-M. Psychological approaches to the study of saving. (En något omarbetad och utvidgad version av Studier i ekonomisk psykologi: 34.) Urbana, Ill.: University of Illinois, Bureau of Economic and Business Research, 1970. 114 s. En begränsad upplaga försäljes genom EFI.

55. Sandell, R.G. Situational factors in choice behavior - four research papers. EFI, 1970. Stencil, 45 s.

56. Sandell, R.G. The effect of instruction perspective, detail, and medium on learning and generalization of a discriminative habit. EFI, 1967. Stencil, 16 s.

57. Hellström, B. & Savelius, H. Preferenser för löneökning eller arbetstidsförkortning. En intervjuundersökning bland SIF-tjänstemän i Stockholm våren 1969. EFI, 1970. Stencil, 37 s.

58. Wärneryd, K.-E. Can results from psychological laboratory experiments be generalized to situations outside the laboratory? Ingår i Lindblom, P. (ed.): Theory and methods in behavioural sciences, s. 73-89. Stockholm: Svenska Bokförlaget/Norstedts, 1970. Särtryck kan erhållas från EFI.

59. Jürss, P. & Terwander, G. Löntagares kunskap om och inställning till det svenska skattesystemet. En intervjuundersökning bland metallarbetare och industritjänstemän i Stockholm. EFI, 1970. Stencil, 65 s.

60. Nowak, K. & Wärneryd, K.-E. Kommunikationsforskning och kommunikationslära - två artiklar. EFI, 1970. Stencil, 16 s. (Artiklar tidigare införda i Den Svenska Marknaden, 1965, nr 5, samt i Ekonomen, 1969, nr 8.)

61. Nowak, K. Resistance against persuasion. EFI, 1970. Stencil, 18 s.

62. Sandell, R.G., Nowak, K. & Ölander, F. Beteende: en funktion av attityder eller situationer? EFI, 1970. Stencil, 58 s. (Tre artiklar tidigare införda i Markedskommunikasjon, 1967, 4 (2), 67-85; 1968, 5 (1), 30-39; 1969, 6 (2), 28-46.)

63. Lindhoff, H. Kommersiell manipulation - ett försök till analys. Ingår i "I eller emot samhällets intresse " - åtta artiklar, s. 17-39, Stockholm: Beckmans, 1970. (Konsultkollegiets kompendier, nr 4.)

64. Cagnell, B. & Cagnell, L. Opinionsledarskap och innovationsbenägenhet för mode. En empirisk studie av 100 flickor i Stockholm våren 1969. EFI, 1970. Stencil, 69 s.

65. Wärneryd, K.-E., Carlman, B., Carlzon, J., Cassel, U., Nowak, K. & Thorslund, S. Reklam och uppmärksamhet. Några artiklar och uppsatser. EFI, 1970. Stencil, 86 s.

66. Lindholm, C. & Nyman, H. Effekten av direktreklam under olika betingelser. EFI, 1970. Stencil, 37 s.

67. Larsson, G. & Looström, L. Dagspressen som gate-keeper. EFI, 1970. Stencil, 48 s.

68. Håkansson, M. & Högberg, P. Informativ reklam kontra konsumentupplysning: Studium av kommunikatoreffekter i en valsituation. EFI, 1970. Stencil, 42 s.

69. Stolt, B. The congruity principle and the effect of relation strength. EFI, 1971. Stencil, 105 s.

70. Ölander, F. Är konsumenterna olika? Ingår i Rasmussen, A. m.fl.: Det segmenterede salg, s. 53-65. København: Nyt Nordisk Forlag/Arnold Busck, 1970. Särtryck kan erhållas från EFI.

71. Wärneryd, K.-E. Kommunikationsvägarna i samhället. Ingår i Mattsson, L.-G. (red.): Människor och företag i kommunikationssamhället, s. 1-14. Stockholm: Prisma/EFI, 1971.

72. Nowak, K. Från informationsteknik till kommunikationsteknik. Ingår i Mattsson, L.-G. (red.): Människor och företag i kommunikationssamhället, s. 15-40. Stockholm: Prisma/EFI, 1971.

73. Lundberg, D. Massmedia och samhällsbevakningen. Ingår i Mattsson, L.-G. (red.): Människor och företag i kommunikationssamhället, s. 41-68. Stockholm: Prisma/EFI, 1971.

74. Lindhoff, H. & Ölander, F. Konsumenternas inflytande på företagens produktutveckling. Ingår i Mattsson, L.-G. (red.): Människor och företag i kommunikationssamhället, s. 147-193. Stockholm: Prisma/EFI, 1971.

75. Wärneryd, K.-E. Om forskningskommunikation. Ingår i Mattsson, L.-G. (red.): Människor och företag i kommunikationssamhället, s. 265-290. Stockholm: Prisma/EFI, 1971.

76. Björkman, J. Kortsiktiga effekter av trafikinformation. Stockholm: EFI, 1971. 416 s.

77. Tuveson, J. & Wöhrmann-Hill, N. Strängaspelet - några faktorer som påverkar det enskilda sparandet. EFI, 1971. Stencil, 55 s.

78. Lundberg, D. & Hultén, O. Massmediaforskning enligt användningsmodellen. EFI, 1971. Stencil, 60 s.

79. Gärdborn, I. Ett experiment med Delfimetoden. EFI, 1971. Stencil, 119 s.

80. Enqvist, M. & Jacobson, C. Uppfattningar om börsföretagens information till aktiemarknaden. EFI, 1971. Stencil, 45 s.

81. Nowak, K. TV-producenters och publikens bedömningar av TV-program - en explorativ studie. EFI, 1971. Stencil, 26 s. (Ingår också i appendix till Studier i ekonomisk psykologi: 82.)

82. Nowak, K. The psychological study of mass communication effects: On the validity of laborating experiments and an attempt to improve ecological validity. EFI, 1972. Stencil, 189 s. + separat appendix.

83. Fjaestad, B. & Nowak, K. Massmedia och företagen. Stockholm: SNS/EFI, 1971. 395 s.