Studies and experiments towards a computer interview acceptable to the naive user

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/10499

Publisher: © Vassilis Spiliotopoulos

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
ABSTRACT

This thesis is concerned with concepts and procedures involved in the development of programs by which computers may interview people.

Based on the hypothesis that the style of the questions is very important in a computer interview, an analysis of four medical interviewing programs was attempted. Two variables emerged from this analysis namely Encouragement (Encouraging Phrases) and Chattiness (Chatty Phrases).

Six experiments were then conducted to test the usefulness of the above two variables in a computer interview.

The results of the six experiments were as follows:

a) The use of random Encouragement and Chattiness seems to have a significant effect on interviewee's acceptance of a computer interview.

b) There appears to be an optimum for the amount of Encouragement and Chattiness used in a computer interview.

c) Random Chattiness seems to have no effect on interviewee's acceptance of a human interview, while random Encouragement might have a negative effect.

d) For a computer interview random patterns of Encouragement and Chattiness seem as good as more planned patterns of the two variables.

e) The accuracy of the derived information and the interviewees' acceptance of a computer interview on a general subject may be as good as that for a human interview.

From these results the original hypothesis is seen to be too limited and in its place a conceptual analysis of the essential features of computer based interview programming is developed.
STUDIES AND EXPERIMENTS TOWARDS A COMPUTER INTERVIEW
ACCEPTABLE TO THE NAIVE USER

Vassilis Spiliotopoulos

Department of Human Sciences
University of Technology
Loughborough
Leicestershire

Ph.D. Thesis
November 1977
ACKNOWLEDGEMENTS

I wish to thank everyone who helped me in any way to complete this thesis.

In particular I want to thank my supervisor Prof. B. Shackel for his guidance and encouragement through this work, Mr. Tim Folkard and Mr. Dick Penn of the National Physical Laboratory (N.P.L.) team, and Mr. Rudi Dallos, all within the HUSAT research group, for their constructive comments at the beginning of this work.

I would also thank N.P.L. for supporting this research, Dr. C. Evans of N.P.L. and his team for supplying me eagerly with every information and help concerning the medical interviews and Mr. N. Bevan for kindly lending me his programs, which I used for some of my experiments that were run in the PDP 11/10 mini-computer.
CONTENTS

ABSTRACT 1
DEFINITIONS 2
ABBREVIATIONS 3

CHAPTER 1: INTRODUCTION AND AIMS 4

1.1 Medical Interviewing Programs 4
  1.1.1 Medical History-Taking Programs 5
  1.1.2 The work of N.P.L. London 9
1.2 Computer Interviews in General 13
1.3 Aims 16

CHAPTER 2: AN ANALYSIS OF FOUR MEDICAL INTERVIEWING PROGRAMS 18

2.1 Structure of the Flow-charts of the Programs 19
2.2 The Introduction of the Programs 19
2.3 Analysis of Simple and Straightforward Questions (S.S.Q.s) 22
  2.3.1 Standard forms of Questions 22
  2.3.2 Elements of Surface Structure of S.S.Q.s 24
  2.3.3 The Problem of Semantics 25
  2.3.4 Transformation Rules 25
2.4 Analysis of Linguistic Style Elements (L.S.E.) 25
  2.4.1 Phrases preceding the question 26
  2.4.2 Questions differing from S.S.Q.s 32
  2.4.3 Phrases following the question 37
2.5 Patterns of L.S.E. 40
  2.5.1 Patterns of Encouraging Phrases (E.P.s) 40
  2.5.2 Patterns of the rest of L.S.E. (apart from E.P.s) 42
2.6 Correlation between L.S.E. and the Meaning of the Questions 43
2.7 Discussion on the Analysis 47

CHAPTER 3: RATIONALE: FROM THE ANALYSIS AND PLANS OF THE EXPERIMENTAL PROGRAMME 52
3.1 Rationale from Chapter 2 regarding Research Aims 52
3.2 General Plan and Experimental Programme 53
3.3 Methods
  3.3.1 Material 54
  3.3.2 Variables 54
  3.3.3 Subjects 55
  3.3.4 The Computer System 55
  3.3.5 The Man-Computer and the Man-Man Interaction 55
  3.3.6 The Semantic Differential Method 56
  3.3.7 Analysis of Results 56

CHAPTER 4: AN EXPERIMENT TO DEVISE A METHOD FOR MEASURING PEOPLE’S ATTITUDE 57
4.1 Aims 57
4.2 Material 57
4.3 Variables 58
4.4 Subjects 59
4.5 The Semantic Differential Method 59
4.6 The Computer System 61
4.7 The Man-Computer Interaction 62
4.8 The Factor Analysis - Results with the Semantic Differential Method 63
4.9 Further Analysis of Results - The Questionnaire Versions 66
4.10 Conclusions 73

CHAPTER 5: AN EXPERIMENT TO COMPARE SEVERAL RANDOM PATTERNS OF E.P.s 74
5.1 Aims 74
5.2 Material 74
5.3 Variables 74
5.4 Subjects 75
5.5 The Computer System 75
5.6 The Man-Computer Interaction 76
5.7 The Semantic Differential Method 77
5.8 Analysis of Results 78
5.9 Conclusions 82

CHAPTER 6: AN EXPERIMENT TO COMPARE SEVERAL RANDOM PATTERNS OF L.S.E. IN A HUMAN INTERVIEW 84
6.1 Aims 84
6.2 Material 84
6.3 Variables 84
6.4 Subjects 85
6.5 The Man-Man Interaction 85
6.6 The Semantic Differential Method 91
6.7 Analysis of Results 91
6.8 Conclusions 92

CHAPTER 7: AN EXPERIMENT TO COMPARE RANDOM VERSUS NON-RANDOM PATTERNS OF L.S.E. IN A COMPUTER INTERVIEW 99
7.1 Aims 99
7.2 Material 99
7.3 Variables 99
7.4 Subjects 101
7.5 The Computer System 101
7.6 The Man-Computer Interaction 101
7.7 The Semantic Differential Method 102
7.8 Analysis of Ratings 103
7.9 Conclusions 112

CHAPTER 8: AN EXPERIMENT TO COMPARE PEOPLE'S ATTITUDE TOWARDS COMPUTER AND HUMAN INTERVIEWS 113
8.1 Aims 113
8.2 Material 113
8.3 Variables 113
8.4 Subjects 114
8.5 The Computer System 114
8.6 The Man-Computer Interaction versus Man-Man Interaction 114
8.7 The Semantic Differential Method 115
8.8 Analysis of Results 115
8.9 Conclusions 119

CHAPTER 9: COMBINATION AND REVIEW OF RESULTS FROM THE FIVE EXPERIMENTS AND GENERAL DISCUSSION THEREON 120
9.1 Summary of Thesis so far 120
9.2 The Method for Measurement of Acceptability of Interviews 122
9.3 The influence of Encouraging and Chatty Phrases 122
9.4 Comparison of Different Patterns of Encouraging and Chatty Phrases 123
CHAPTER 10: MEASUREMENTS OF REPRODUCABILITY AND ACCURACY IN A COMPUTER INTERVIEW

10.1 Aims
10.2 Material
10.3 Variables
10.4 Subjects
10.5 The Computer System
10.6 Man-Computer Interaction versus Man-Man Interaction
10.7 Analysis of Results
10.8 Reproducability Results
10.10 Conclusions

CHAPTER 11: DISCUSSING THE RESULTS OF THE EXPERIMENT ON REPRODUCABILITY AND ACCURACY OF INFORMATION

CHAPTER 12: THE CONSTRUCTION OF THE COMPUTER SYSTEM

12.1 Rules for the Construction of the Computer System
12.2 Rules for storing the E.P.s and C.P.s in the System
12.3 Rules for the construction of the first part (stage 1) of the Computer System
12.4 Rules for the construction of the second part (stage 2) of the Computer System

CHAPTER 13: GENERAL DISCUSSION

13.1 Summary of the Analysis
13.2 The Experimental Methods
13.3 The Results
13.4 The Computer Software System
13.5 Suggestions for the Maintenance and further Perfection of the Computer System
13.6 Limitations of the Research and the Results
13.7 A Revised Formulation of Computer Interviewing
13.8 Have the Original Aims been fulfilled and to What Degree?
13.9 Conclusions

CHAPTER 14: CONCLUSIONS

REFERENCES

APPENDIX 1 - FIGURES

APPENDIX 2 - THE COMPUTER SYSTEM

APPENDIX 3 - INTRODUCTION, E.P.s, C.P.s
ABSTRACT

This thesis is concerned with concepts and procedures involved in the development of programs by which computers may interview people.

Based on the hypothesis that the style of the questions is very important in a computer interview, an analysis of four medical interviewing programs was attempted. Two variables emerged from this analysis namely Encouragement (Encouraging Phrases) and Chattiness (Chatty Phrases).

Six experiments were then conducted to test the usefulness of the above two variables in a computer interview.

The results of the six experiments were as follows:

a) The use of random Encouragement and Chattiness seems to have a significant effect on interviewee's acceptance of a computer interview.
b) There appears to be an optimum for the amount of Encouragement and Chattiness used in a computer interview.
c) Random Chattiness seems to have no effect on interviewee's acceptance of a human interview, while random Encouragement might have a negative effect.
d) For a computer interview random patterns of Encouragement and Chattiness seem as good as more planned patterns of the two variables.
e) The accuracy of the derived information and the interviewees' acceptance of a computer interview on a general subject may be as good as that for a human interview.

From these results the original hypothesis is seen to be too limited and in its place a conceptual analysis of the essential features of computer based interview programming is developed.
DEFINITIONS

The following terms are defined at this stage, because they appear frequently in the text starting from the very beginning.

a) Simple Straight Questions (abbreviated S.S.Q.s).
S.S.Q.s are the simplest and shortest piece of a text to convey an intended message belonging to the class of valid English interrogative sentences.

b) Linguistic Style Elements (L.S.E.)
L.S.E. are all the phrases, differing from S.S.Q.s, by which the interviewing program conveys messages to the user.

c) Surface Structure of a Sentence.
This is a proper bracketting of the linear, temporally given sequence of elements, with the paired brackets labelled by category names (that is a labelled tree diagram, with such categories as Sentences, Noun Phrase, Verb Phrases, Noun and a small number of others serving as labels). An example of how a sentence can be split into its surface structure elements is given in Figure 1.

d) The deep structure of a sentence is in general not identical to its surface structure, but is a much more abstract representation of grammatical relations and syntactic organisation.

The definitions a) and b) appear for the first time in this study. The definitions c) and d) are as in N. Chomsky's 'Topics in the Theory of Generative Grammar'. (Chomsky, 1966).

An exception of a) could happen for a context dependent question. For example, a S.S.Q. following a sequence of related questions could appear without a verb.
ABBREVIATIONS

Most of the following abbreviations are used in Figures 2, 3, 4 and 5 showing the flow-charts of the programs. Some of them are also used throughout the text.

Apart from the two abbreviations S.S.Q.s = simple straight questions and L.S.E. = linguistic style elements, which have already been given, the following abbreviations are used:

I = Introduction of a program
Pr.P. = Pre-phrase, that is a phrase preceding a question
E.P. = Encouraging Phrase
Pr.P3. = Freely attached or detached pre-phrase
Pr.P4. = Other types of pre-phrase
S.F. = Standard forms of questions
I.T.Q. = If-then Questions
Q.I.P. = Question with included (or embodied) phrase
D.Q. = Descriptive question
Po.P. = Post-phrase, that is a phrase following a question
Po.P1. = Post-phrase explaining part of the question
Po.P2. = Post-phrase giving directions to the user
C.P. = Chatty Phrase
1. INTRODUCTION AND AIMS

This thesis describes research studies on man-computer interviewing. These studies were part of a collaborative research programme with the National Physical Laboratory (N.P.L.) London on man-computer interaction problems, with the research in this thesis being entirely the author's work (under the customary guidance and supervision).

The first part of the thesis reports on an analysis of four medical history-taking interviewing programs, representative of the work of Dr. Evans and his team at the N.P.L. This part is comprised in Chapter 2. The second part consists of six experiments aiming to test the usefulness of the findings of the previous analysis for a computer interview of a general subject. This part is contained in Chapters 3-11. Chapter 12 consists of rules for the construction of the proposed computer system. The main results and the essential issues of this research are then discussed in Chapter 13, while Chapter 14 consists of the conclusions of the work described so far.

The rest of this introductory Chapter 1 is split into two parts, the former describing the development of computer medical interviews (section 1.1), the latter reporting on computer interviews in general (section 1.2).

Moreover the computer medical interviewing programs of N.P.L. are described in more detail in a separate section of the computer medical interviews (section 1.1.2), because of their close connection to this project.

1.1 Medical Interviewing Programs

During the last two decades the old image of the computer as just a fast arithmetic calculator has changed dramatically, after the realisation of its potential as a means of control and communication.
One field of relatively recent origin, about ten years ago, among those where the computer is used as a means of communication, is man-computer interviewing with the computer simulating the interviewer.

Among the several possible areas of man-computer interviewing that of medical history-taking has attracted most of the attention and effort of the researchers. There are two main reasons for this concentration of research on medical history-taking interviewing programs:

The first reason is the obvious tremendous importance of medicine for human life.
The second reason is the fact that most of the illnesses can be identified through a number of well-defined symptoms, and therefore a computerized questionnaire elicitng information about those symptoms can easily be constructed.

1.1.1 Medical History-Taking Programs

The aim of medical history-taking programs is to assist the doctor by handling the routine but important job of taking the medical history of patients, thus saving some of his/her valuable time.

These programs usually present the questions through a teletype or a V.D.U. in natural language, while the patient responds by choosing among a limited number of possible responses. They are normally evaluated in terms of their acceptability to the patient, the precision and accuracy of the derived information, and the monetary cost.

Their use is still, probably with a few exceptions (e.g. the MUMPS computer system in the Massachusetts General Hospital), on an experimental and research basis rather than on a regular one. The main reasons for this are the lack of standardisation of the techniques, and of the exact criteria for their evaluation, and the problem of their acceptability to the clinicians.
Yet there are several successful history-taking programs, which therefore encourage both further research in this field and the possibility of their use on a regular basis in the near future. This is indicated by the characteristic examples of computer history-taking interview programs which are described below. (The following examples give us also an idea of the evolution of the computer interviewing programs and especially of their evaluation criteria from 1965 till now.)

One of the first computer history-taking programs was operational at Wisconsin University by as early as the spring of 1965 (Slack et al., 1968). The authors considered the new process as an improvement over the self-administered questionnaires printed on punched data-processing cards, the latter probably being the first attempt towards the automation of the medical history-taking process, for the following reasons:

1) it eliminated the need for intermediary manipulation of data between the source (the patients) and the computer;
2) it enabled rapid, complex branching and a responsive dialogue between patient and computer; and
3) it permitted control over the interview in a physician-like way.

Yet the authors mentioned no measurement or evaluation based on the above three criteria.

In 1968 the Automated Medical History (AMH) system at the Mayo Clinic was reported by Mayne et al. (1968). The AMH system was evaluated among others for acceptability to the interviewees and accuracy of the derived information. The former, acceptability, was based on open-ended answers of 154 patients to direct questions about computer administration of medical questions, and on the patient's general remarks on the system; these replies were favourable to the system (140 favourable, 14 unfavourable, 3 refusals to participate because of an adverse reaction to computer administration). The latter,
accuracy, was considered satisfactory because the system recorded 96% of the symptoms recorded by the physicians (no measure of the agreement between physicians was given by the authors).

In the period 1970-1974 a number of computer history-taking programs was reported by Evans and his group at the N.P.L. These programs were evaluated for acceptability to the interviewees and accuracy and precision of the derived information. The first was based on people's answers to direct questions and on respondent's comments, and was reported to be very favourable to the computer (e.g. Evans et al., 1973). The second was based on the relative agreement of two consultants and the computer, after interviewing the same 72 patients on the same subject, and on a statistical estimation of their errors, and was found to be as precise and accurate as the consultants' interviews (Evans et al., 1971; Card et al., 1974).

In 1977 two studies in computer medical interrogation were reported in a paper by Lucas et al. (1977). The two programs described were evaluated for acceptability to the patient, for accuracy and precision of the derived information, and for monetary cost, by using 75 and 36 patients respectively. The patients' acceptance was measured by using recognised psychological scaling techniques (a modification of Edwards' 'scale discrimination method' and the Semantic Differential), and resulted in patients' ratings of computer interrogation as favourable as those of doctor interrogation. Using statistical models to estimate the errors of the computer and a number of consultants (3 for the first study, 2 for the second one), and the relative agreement between them, the accuracy and precision of the computer was found comparable to that of specialist physicians. (The consultants and computer were interviewing the patients on the same subject, as in the case of Dr. Evans's programs.) As to monetary cost the authors commented that it could be cheaper than that of a consultant, if a mini-computer serving several interrogation terminals simultaneously was used, while the new technology of microprocessors would make it even cheaper.
The above are but few examples of computer history-taking programs successfully demonstrated on actual patients. The use of such programs with varying acceptability to patients has been reported in most areas of medical science including the following: Alcoholism (Reich et al., 1975; Lucas et al., 1977), Ante-Natal History (Fuller, in print; Pearlman et al., 1973), Chest Diseases (Evans et al., 1974), Child Psychiatry (Coddington et al., 1972), Dyspepsia (Lucas et al., 1977), Emergency Room Patients (Greist et al., 1973), Eye Examination (Marg et al., 1972), Family Personal and Social History (Eaton et al., 1970), Functional Headache (Stead et al., 1972), Gastric Ulcer (Evans et al., 1971), Hysteria (Woodruff et al., 1973), Preventive Medicine Suicide Risk Prediction (Greist et al., 1973), Psychiatry in General (Greist et al., 1973), Psycho-Sexual History (Martin, in print; Slack, 1971), Venereal Disease (Vancura, 1975), Urological Diseases (Pearlman et al., 1972), Uterine Cancer (Peckham et al., 1967) etc.

Yet all these research studies have not led to as much practical applications as might be expected, because of problems of acceptability to patients and to clinicians. Therefore it seems that much more research is required on the human factors involved in a successful man-computer interaction, and that these human factors should be taken seriously into account during the design and implementation of computer-based interview programming systems. The evolution of the computer medical interviewing programs in the U.S.A. during 1965 to 1975 can be viewed as a clear indication of the above fact. Accordingly in U.S.A., because they tried to rush in the applications of the medical interviewing programs, without much research and understanding of the human factors involved, they did not have good results. This resulted in less enthusiasm about the idea of medical interviewing programs and in fewer such programs in the period 1970-1975 to those of the period 1965-1970. (The above was assessed by the author through personal communication with participants of the Third International Conference on Computing in the Humanities, held in Waterloo, Canada, 2-5 August, 1977.)
The problem of acceptability to clinicians, which is very important if the computers are to be used on a regular basis for patient interrogation, has been underlined by a number of researchers (e.g. Mayne, 1972; Klein et al. 1975; Lucas et al. 1977) and by the Committee of Public Accounts (Committee of Public Accounts, 1976).

For the problem of the adaptation of the computer system to the individual patient a number of suggestions have been made by Lucas. These suggestions include the adaptation of the speed of presenting the questions and the number of responses available to the patient, depending on the age and intelligence of patients, the latter being decided by the patient's response times to the introductory question (Lucas, 1974).

The realisation of the need for better understanding of the human factors involved in the interview process, as described above, was the starting point of the research work described in this thesis.

This has further led to the idea of analysing successful medical interviewing programs to aid the understanding of the computer interview process and especially of the linguistic framework of an acceptable interviewing questionnaire program. Towards this direction the work of N.P.L. London on medical interviewing programs, which is described below, was chosen to be analysed. (Detailed reasons for this choice are given at the beginning of Chapter 2.)

1.1.2 The Work of N.P.L. London

During the period 1970-1974 a major part of the research program of the Man-Computer Interaction section of N.P.L., under Dr. Evans, was devoted to the development of computer history-taking interview programs. The method used for the construction of these programs was as follows.
First one or more experienced clinicians were consulted to define the exact symptoms of the disease and the useful information about each symptom. The relevant questionnaire was then constructed aiming to elicit the needed information concerning the symptoms of the disease, and the computer program was written incorporating this questionnaire. During a few trials in the subsequent pilot study the comprehensiveness of the questions was tested and altered accordingly, if this was judged to be necessary.

The computer programs were written in the Basic computer language and the questions were presented either on the teletype or on a V.D.U., usually at a rate of 10 characters per second. During the interaction the patients were responding at their own pace by pushing one of the three buttons ('YES', 'NO', '?') of a mask clipped over the teletype keyboard. This design simplified the task of the respondent and minimized the probability of an error.

Among the several aspects of the evaluation of the medical history-taking programs major emphasis was given to their acceptability to the patients. This was thought to be improved considerably by the style of the programs, which put the questions in a friendly, chatty way, closely following the characteristics of an informal history-taking (Evans et al. 1971).

These natural and friendly computer interviews were the result of a deliberate effort of an experienced psychologist (Dr. Evans) who, having realised the importance of this factor to interviewee's acceptance, attended a series of preliminary diagnostic sessions, conducted by various consultants, to collect information about their style and manner. Then he tried to build the computer 'personality', as expressed by the form of its questions, as an amalgam of the best style and manner of the specialists involved in the interviews he attended, adapted somewhat to 'read well' on the computer presentation (Evans et al. 1971).
The four most characteristic medical interviewing programs of Dr. Evans and his team in N.P.L. have to do with PREGNANCY, GASTRIC ULCER, SEX and CHEST complaints correspondingly. Their acceptability to the patients was assessed by interviewing the patients after the computer interview (the assessment interview consisting mainly of direct questions), and by the spontaneous comments of the interviewees.

The program PREGNANCY was tested for acceptability to the interviewees by a sample of 34 patients. Out of those 34 patients only two were against the use of the system, because they would prefer a more 'human touch', they did not like a machine to direct the discussion, they would prefer to ask questions also and to answer before the end of the questions. Eight more patients, i.e. ten altogether (29.4%) would prefer a doctor/nurse interview to that of the computer. Out of the remaining 24 patients, thirteen (38.2%) had no preference between a doctor/nurse and a computer interview, while eleven patients (32.4%) preferred the computer interview.

Reasons given by the patients in favour of the computer interview were the following:
They could answer the questions at their own pace; they did not waste anyone's time; they were not 'looked at' by anyone; and they felt free to answer questions of a personal manner (Fuller, 1977).

Two main negative comments were made by the patients for the computer interview. One was the restriction of the YES/NO answers, the other that there was no way to correct their mistakes. Among the spontaneous comments of the patients phrases like the following appeared: 'Marvellous idea', 'fantastic', 'very chatty and polite', 'great idea', 'good fun', 'a very good idea', 'very simple and straightforward', 'quite an experience', 'wonderful novelty', etc.
Analogous results to those of the program PREGNANCY were reported for the other three programs.

The acceptance of the program GASTRIC ULCER, as assessed by interviewing 75 patients after the computer interview, was as follows (Card et al. 1974). To the question 'Did you find the experiment interesting?' only one out of the 75 patients, described by the assistant as very slow and asking for help frequently, gave a negative reply (1.3%). The question 'Did you find it frightening or unpleasant?' elicited a 100% negative response. In parallel to the last question, both the questions 'Did the computer seem to pay attention to your answers?' and 'Did you feel that there was something talking to you?' elicited a 100% positive response, suggesting an excellent rapport between computer and patient. To the question 'Would you mind if computers like this were used a lot in hospitals?' 94% gave a negative reply and similarly 94% answered 'no' to the following question 'Do you think that other people would mind?'. Moreover, among the total of the interviewed patients, 10% stated that they would definitely prefer the computer interview to that of a doctor. Among the reasons given in favour of the computer interview was that it was 'less embarrassing' than the doctor interview, and that they 'felt more free to speak'. Some patients stated that they felt 'less nervous', and that the computer gave them 'more time to think than the doctors'. Other remarked on how 'polite' it was.

The main negative reason given was the constraint some patients (40%) felt by the YES/NO alternatives. The majority of the 75 respondents (60%) made very favourable comments like: 'Very good', 'very interesting', etc.

The acceptance of the program CHEST was assessed by two studies (Evans et al. 1973, 1974). In the first not very quantitative one, 50 patients interviewed after the computer session showed a very favourable attitude towards the computer interview, reflected in the following remarks: 1) They stated that they found the computer friendly and polite. 2) Many stated that they felt able to be more truthful to it and not withhold details. 3) The computer was judged to be clear and understandable.
4) Patients appreciated being able to 'take their time'.
5) Patients reported appreciating the fact that it has 'understood' their responses, comparing this practice favourably with the non-committal approach of many doctors.

In the second study 20 patients were interviewed after the computer interview. Among these 20 patients, only three (15%) stated that they preferred an interview by doctor to that by computer, while five (25%) stated that they rather preferred the computer interview. The remaining 60% of the patients stated that they would not mind to have either a computer or a doctor interview.

Reasons given by the patients in favour of the computer included that they were in no hurry to answer, it was ideal for dealing with embarrassing questions, they did not feel frightened with the machine as they sometimes do with doctors and that they did not feel that they wasted the doctor's time this way. The main negative reasons given by the patients were the constraints of the YES/NO answers, and the lack of provision of a deletion key. The majority of the interviewees made very favourable spontaneous comments like: 'pretty good', 'interesting', 'personal', 'polite', 'a very good idea', 'tremendous machine', etc.

1.2 Computer Interviews in General

It has been suggested for medical interviews that the full potential of the information elicited by computer interrogation cannot be exploited, unless this information is handled by statistical decision theory to calculate decisions (Card, 1967; Knill Jones et al. 1973).

Further, most of the non-medical computer interviewing programs are incorporated in larger systems, which can combine the information elicited by the interrogation, with information stored in the computer, to calculate probabilities and decisions. These results are then either considered by a decision-maker other than the interviewee, or are presented back to the interviewee in order to assist him/her to take his/her own decisions. The
former category consists mainly of computer administered tests, while the latter category consists of computer counselling and teaching systems. However, compared with medical interviews, rather few publications have been made on computer non-medical interviews. (Here we do not refer to the teaching systems using C.A.I.; these, though they include a kind of a computer interview, consist essentially of a specialist type of situation and communication which is not very similar or relevant and about which much has already been published.

Several studies have indicated the usefulness and validity of computer administered psychological tests.

On the reliability of these tests, Griffith and Elwood (1969) reported a test-retest reliability of 0.99 for the automated Wechsler Adult Intelligence Scale.

On the computer presentation and scoring versus the conventional one, Hedl et al. (1971) reported a correlation of 0.75 between scores of a computer-based and a conventional Slossom Intelligence Test. Hitti et al. (1971) found no significant difference between a computer-based and a conventional Progressive Matrices Test, and Lushene et al. (1974) reported a very high correlation between a completely computerised and a conventional Minnesota Multiphasic Personality Inventory.

On the effects of interaction, the results of several studies indicated that man-machine interaction may be less stressful and biased than face-to-face interaction, which is especially useful in certain sensitive social situations (e.g. Smith, 1963; Johnson et al. 1973). These studies show that this less threatening, more natural image may be extremely important when dealing with certain types of people who are highly apprehensive when they are examined. The lack of human biases may also increase the reliability of the test. In fact subjects often find working with a computer terminal a pleasant experience, and therefore tend to generalise this interest in the method to an interest in the particular task (Hedl et al. 1971; Johnson et al. 1971).
Recently computer systems have made their appearance to administer not a single test, but a battery of standard measures of intelligence and personality (e.g. Cory, 1974). The very few evaluation studies that have been conducted indicate that these batteries seem adequate and enormously time saving (e.g. Paitich, 1973).

Burli (1975) reviewed the literature on computerised test administration and concluded, among others, that the efficiency of computerised tests would be highly improved if techniques of branching or individually tailored testing were used. Results so far demonstrate that branched testing produces a substantial reduction of test length and testing time, while not reducing the level of validity (Burli, 1975). Yet, very few systems have been reported to be tailored to the ability of the examinee (e.g. Bayroff et al., 1974).

The possible usefulness of computer assisted counselling has also been indicated by several studies. Systems capable of interacting with individual users for career exploration have been suggested and developed with good results (e.g. Harris, 1968; Loughary, 1970; Peter & Donnan, 1972).

In an evaluation study of computer assisted vocational choice, compared with traditional vocational counselling, Melhus (1971) suggested that both these modes of counselling are equally effective for willing and well-motivated people, while for less willing or well-motivated people the human counsellors are more effective. In another evaluation study Pilato et al. (1975) have shown that Computer Mediated Vocational Guidance has a positive effect on the appropriateness of vocational preference.

Price (1974) reported a study on a direct comparison between human counsellor and computer effectiveness in helping students to select courses. The result of this study is that no significant differences were found in grades received and number of courses changed, between the students using counsellor-assisted and those using computer-assisted procedures. Moreover, this study suggested that when the student's goals, interests and abilities
are recognised and contribute to the decisions being made, then the student feels that he/she is being treated as a person, regardless of whether he/she is interacting with a computer or a human being. Therefore this study indicates the need for the dissemination of occupational information on a highly personalised basis, as was argued long ago by Brayfield (1951).

1.3 Aims

The importance of the style of the questions for the interviewees' acceptance of a computer interview has been underlined by several researchers (e.g. Evans et al., 1971 & 1973; Slack et al., 1968). Yet rather few researchers have used the style of the questions for improving the acceptability of computer interview to the interviewees.

The phrases producing this acceptable style of the questions were called, for this study, Linguistic Style Elements (L.S.E.), being distinguished from the Simple and Straightforward Questions (S.S.Q.s). (The definitions of L.S.E. and S.S.Q.s are given in page 2, while both are examined in detail in Chapter 2.) The idea that the L.S.E. have a significant effect on interviewee's acceptance of computer interviews was a hypothesis underlying this work.

Based on this hypothesis the first aim of the research was to investigate the L.S.E. in an acceptable computer interview by studying the patterns which these phrases follow among the questions and their relation to the meaning of the questions. This investigation would improve our understanding of an important feature of an acceptable interview, i.e. the linguistic content of it, and would hopefully produce rules enabling the construction of a computer system, which would use the linguistic element to increase the interview's acceptability to the interviewees.
The second aim of this research was to construct and evaluate a computer software system, based on the findings from the analysis and experiments, to carry out interviews with a high probability of being acceptable. In particular the proposed computer system would have the following two properties:

1) After introducing the L.S.E. into the system and feeding it with a questionnaire of S.S.Q.s together with the logical structure connecting them (i.e. the number of the next question after each possible answer among a discrete number of choices), the system would be able to modify the text to become more natural and friendly.

2) Such a system could also produce the interviewing program, which in turn could carry out the interview at a later stage using the revised questionnaire.

In this way a further step could be achieved towards the automation of medical interviews, through the replacement of the programmer and perhaps towards the automation of the interview in general, while at the same time satisfying the need of people for natural and friendly interaction. The reasons supporting the general applicability of the system for a computer interview of any kind are the following:

a) the medical history-taking process is just a special kind of an interview, and

b) most of the L.S.E. would be expected to be valid for an interview of any kind.

Among the L.S.E. those more suitable for the proposed system are the ones independent of the meaning of the S.S.Q.s which could in fact carry out the interview even if no L.S.E.s were used; this is because such independent L.S.E. can be handled more easily in a computer program, and inserted by algorithm among the S.S.Q.s. Therefore the third major aim of this research was to test the effect of the context-free L.S.E. on interviewee's acceptance of computer interviews and to examine the possibility of using the linguistic factor to improve computer interviews without entering the full complexity of the semantics of the language.
2. **AN ANALYSIS OF FOUR MEDICAL INTERVIEWING PROGRAMS**

To fulfil the first aim of the research, and to find out useful elements for the construction of the proposed software system, an analysis of four medical history-taking interviewing programs was attempted. These four programs were written by Dr. Evans and his team at N.P.L. and have to do with PREGNANCY, GASTRIC ULCER, SEX and CHEST correspondingly.

The four programs mentioned above were chosen for this analysis mainly for two reasons. First, because they had been shown to have a high degree of acceptability to the interviewees when tested on actual patients, as already described in Chapter 1. Second, because one of the main reasons for their degree of acceptability to the interviewees was the 'friendly' and 'chatty' way in which they carried out the interview. The latter was clearly indicated by both the people who wrote the programs and the patients' comments after the computer interview.

It was hypothesised that the L.S.E. derived from the proposed analysis would enable the computer system to carry out 'natural' and 'friendly' interviews, should they be proved to be adequate for this purpose.

Yet, there would be one more minor semantic problem. This is how to provide for the further explanation of a phrase, if the interviewees could not understand it. One way of tackling this problem would be to find, if possible, standard forms of questions and equivalent versions for each standard form, which would convey the same message but with different wording. Another way would be to provide a supplementary explanation of that part of any question which would probably cause the difficulty.

Consequently, the general analysis to be attempted had the following two aims:
1) To analyse the S.S.Q.s from grammatical, syntactic and semantic points of view, in order (a) to find out, if possible, standard forms of questions, and (b) to identify the several surface structures of the S.S.Q.s, the way they were connected, and thus the possible rules for alternative forms and supplementary explanations.

2) To identify and classify the several L.S.E., to investigate the patterns which they follow among the questions, and to find out how the L.S.E. are related to the meaning of the questions.

Emphasis was given to the second aim, and special weight to the Encouraging Phrases preceding the question with the aim to improve the user's positive attitude towards the system.

2.1 Structure of the Flow-charts of the Programs

The structure of the flow charts of the four programs, more precisely the structure of the flow charts of the questions and phrases as they could be presented by the programs, is in the form of a central trunk and several side branches. These side branches, however, are not free or open-ended, but connected and all return to some point in the main trunk. Therefore it could be said that this structure is linear permitting only closed branches. The only exception is program SEX, which is separated at an early stage into two main trunks corresponding to each of the sexes. Each of these two trunks presents again the type of closed branches mentioned above.

Figures 2, 3, 4 and 5 of the Appendix show the flow charts of the four programs. The abbreviations used in these figures are explained in page 3, while the phrases involved are discussed in detail in section 2.4.

2.2 The Introduction of the Programs

The introduction, that is the phrases used to elicit the necessary initial information from the patient, is separated from the main part which follows. Sometimes it is not easy to
separate the introduction from the following main part. However, there are certain points, independent of the main questions, which are common in most of the four programs, as follows:—

1) Greeting (e.g. 'Hello'): all 4 programs.
2) Outline of what is going to take place and why (e.g. 'This is an experiment in taking details of medical symptoms by a computer. It is designed to help doctors in their work, and to improve medical facilities for the patients as well'): 3 programs out of 4.
3) It asks the user to take part in the interview by a clever persuasive way (e.g. 'We would very much like you to help us with this. If you are prepared to, would you press the button marked 'YES' but if you would rather not, press the button marked 'NO', but we would like you to help us. If you do not understand what to do, press the '?' button. Go ahead now and press one of the three buttons'): 3 programs out of 4.
4) The computer states its identity and its limitations, and consequently the way by which the user can communicate with it (e.g. 'This is a computer talking to you. I can ask you questions but I can only understand simple answers, and you give these answers when you push one of the buttons in front of you. Therefore all the questions can be answered in this way'): all 4 programs.
5) It checks the user's understanding by simply asking him/her if he/she understands (e.g. 'Do you understand everything I've said up to now?'): 3 programs out of 4.
6) It assures the user about the confidentiality of the interview and asks him/her if he/she is happy about that (e.g. 'Please remember that this is all in complete confidence just as with doctors. Only doctors and nurses will see your answers. Are you happy about this? Press one of the buttons'): all 4 programs.
7) It informs the user that he/she has not to rush to answer but he/she can take his/her time and think about a question, if he/she needs to (e.g. 'By the way, don't feel that you have to press the button the very instant that I have finished typing. If you feel you need to think a bit before you answer, or need
to read the question over again, then go ahead and do so. Take your time if you want to': 2 programs out of 4.

8) It reminds the user once more how he/she can communicate with the computer (e.g. 'Remember, however, that I can't talk to you again until you have pressed one of the buttons. I hope that is alright'): 2 programs out of 4.

9) It informs the user that if he/she does not understand exactly a word, because it is scientific, not to worry about it but to declare it by pressing the appropriate button (e.g. 'During the interview I may use medical words, which you may not understand. If you have never heard of these medical words do not worry. I am certain you will recognise the words if you have come across them. If you do not know what they are please answer 'NO''): 1 program out of 4.

There is one more point, which could be considered either as the last part of the introduction or as the first part of the main questions, as follows:-

10) a) It declares that it is going to ask the user a few questions about the subject of the interview, or alternatively, b) it asks the user if he/she has any problem concerning the special subject of the interview (e.g. 'I gather that you have come for help because of some problem with your sexuality or sex life. Is this correct? Push one of the buttons'): 1 program out of 4 for each of a) and b).

Comments on the introduction:
There is a long introduction so that the user may become familiar with the system and feel calm and ready to answer the questions of the following main part.

There are three interaction points in the introduction (numbers 3, 5 and 6 above). For every positive response of the user the system thanks him/her and encourages him/her. The strength of these encouraging phrases seems to decrease from the first to the third interaction point.
Encouraging phrases used at the first interaction point (number 3 above) are: 'Good', 'Thanks very much', 'I hope you will find it interesting', 'We will find it very helpful to us'. At the second (number 5 above): 'Thanks'. At the third interaction point (number 6 above): 'Good'.

Finally, one of the programs has only one interaction point, the third one, and the following encouraging phrase could be considered as equally weighted among the three interaction points of the general introduction: 'I am very pleased to hear that'.

2.3 Analysis of Simple and Straightforward Questions (S.S.Q.s)

2.3.1 Standard forms of questions

A laborious search was made in order to find a small number of general questions for each program, so that each of the questions in the programs belonged to one of these general or standard forms of questions.

The assumption of the possible existence of these standard forms of questions was based on (1) the common form of the questions (YES-NO questions), and on (2) the common contents of the questions (they belong to a medical history-taking questionnaire).

The reason underlying this search for standard forms of questions was to produce a solution to the problem arising when a user could not understand a question. In this situation the question causing the trouble could be replaced by an equivalent one, but with different words, through the standard forms of questions and a look-up table.

The differentiation of one standard form of question from another one was based mainly on the meaning of the main verb (that is the verb following the auxiliary verb of the question or the auxiliary itself if no other verb followed it), or on the connection or not of the main verb with other elements of the sentence (especially objects).
Some of the main verbs were classified into groups of similar meaning, where they occurred with high frequency in the questions examined (e.g. 'Suffer from' group, or verbs of movement or of feeling or of think-how type, etc.). Yet it was not possible to classify all the main verbs according to their meaning, and some were mentioned merely as verbs. In the latter case, the standard forms of questions were based on the transitivity or not of this 'general' verb and, in the former case if the verb was transitive, on the number of objects. For every auxiliary verb, used as a main in the question, a different standard form was created. Finally, a standard form represented the phrases with no verb at all. These phrases depend upon the discourse and usually upon the previous question-answer.

Phrases of the same standard form, on the other hand, differ according to the existence or not of certain grammatical terms, either simple (e.g. adverbs, conjunctions, nouns, pronouns etc.), or compound (e.g. noun phrases, verb phrases, phrases of time, of place, of manner, clauses etc.), and according to the position of these terms in the phrase.

Working this way there were found 7 standard forms of questions for the program PREGNANCY, 6 for the program GASTRIC ULCER, 9 for the program SEX, and 6 for the program CHEST. The total number of different questions for each of the four programs is as follows: Pregnancy 106, Gastric Ulcer 29, Sex 54, Chest 144. However, there is a correlation among the standard forms of questions of the four programs, so that the forms of questions from all the programs can be combined into 14 different standard ones.
Discussing the results of the analysis as far as the standard forms are concerned the following comments can be made:

1) The fewer the standard forms of questions, which we attempt to create, the more complicated their structure (that is more branches and sub-divisions).

2) The involvement of 'open-system' grammatical items (that is items belonging to a class which can be indefinitely extendable e.g. verbs, nouns, etc.) makes it difficult to identify the meaning of these standard forms of questions, but only in special cases.

3) The questions depending on the discourse add further to the difficulty of recognising the meaning of the questions belonging to the corresponding group.

The above arguments are strongly against the idea of using standard forms of questions to solve the problem, which arises when the user does not understand a phrase, by producing an equivalent phrase with different words.

2.3.2 Elements of Surface Structure of S.S.Q.s

On the other hand it could be said, as a result of the above analysis, that certain grammatical terms, simple or compound, can be relatively easily identified in S.S.Q.s together with their syntactical role in the phrase (e.g. the auxiliary, the main verb, the subject (usually you) etc.).

Moreover the connection of the several surface structure elements of S.S.Q.s follows an especially simple pattern at the beginning of the questions. Another favourable fact is that 'Close system' grammatical items (that is items belonging to a class which cannot normally be extended by the creation of additional members, e.g. articles, pronouns, prepositions, conjunctions, etc.) would also ease the identification of compound grammatical elements. For example, a time phrase would start with a time indicator, an adjective clause with a relative pronoun, etc.
2.3.3 The Problem of Semantics

These findings encourage the alternative solution to the problem resulting when a user does not understand a question. That is to analyse the part of the phrase which probably causes the trouble. It seems fair to assume that, in the case of the medical interviews, the patient's not understanding of a phrase is mainly caused because of the medical terms involved. Consequently, a suitable solution would be to test every probable word of the phrase, to see if it exists in a stored dictionary of medical terms (technical terms for an interview of any kind) and if so to state its explanation.

2.3.4 Transformation Rules

It is also possible that the identification of the elements of surface structure of S.S.Q.s could be used for the transformation of questions of other types to S.S.Q. and vice-versa, if certain transformational rules could be found as a result of the relevant analysis which is discussed in the next section.

2.4 Analysis of Linguistic Style Elements (L.S.E.)

The linguistic style elements consist of the following three main parts:

1) Phrases preceding the question.
2) The question itself, if different from S.S.Q.
3) Phrases following the question.

A further classification of each of the 3 parts is as follows:

1.1) Encouraging phrases aiming to improve the user's positive attitude towards the interview (E.P.s).
2) Phrases consisting in part of a question by being included (or embodied) in it (Q.I.P.) - see below.
3) Phrases just preceding a question, to give the user the impression of a natural dialogue, being freely attached to or detached from the question (Pr.P3).
4) The rest of the pre-phrases including all the pre-phrases which do not belong to the 3 groups of pre-phrases described above (Pr.P4).

2.1) If-then questions (I.T.Q.).

2) Questions with included (or embodied) phrases (Q.I.P.).

3) Descriptive questions (D.Q.).

3.1) Post-phrases explaining or analysing further a part of the question (Po.P1).

2) Post-phrases giving directions to the user (Po.P2).

The detailed description of each of the above parts is as follows.

2.4.1 Phrases preceding the question

This section reports on the phrases, which were found to precede the main questions of the four computer interviewing programs, which were analysed.

The phrases preceding the question are classified in four categories, and are presented together with their frequency of occurrence in each of the four interviewing programs.

Furthermore some comments are made on the relation of these phrases to the meaning of the questions, as resulted from preliminary inspection (statistical analysis on this matter is presented in section 2.6).

Among the phrases preceding the question the Encouraging Phrases were also rated according to the amount of encouragement, which a panel of researchers judged them to contain.
A) Encouraging Phrases

An encouraging phrase is defined as a reaction of the system to the user's last answer, which refers to the last question-answer of the interaction.

The purpose of these Encouraging Phrases is, as mentioned above, to increase the user's positive attitude towards the computer interview. Twenty-six different encouraging phrases were found in all four programs. The encouraging phrases as they appear in each of the programs are as follows: (the right upper index is their frequency, being equal to 1 if omitted).

Pregnancy
'Thank you. You are managing to answer my questions very well' / 'Thank you' / 'We have almost finished' / 'Good' / 'Congratulations' / 'Good that's fine' / 'Fine' / 'Good I am pleased to hear that' / 'I am very sorry to hear that'.

Gastric Ulcer
'Thank you' / 'Well you are being very helpful so far. I hope you still find the interview interesting' / 'Thank you. We are getting on very well and I have only one or two more questions to ask you' / 'I see' / 'O.K.' / 'Very well'.

Sex
'O.K.' / 'I see' / 'Alright' / 'Very well' / 'Well that's all very clear' / 'Good' / 'I am so glad' / 'Well that's good' / 'Splendid' / 'Thanks for answering those questions' / 'Fine. I think that's clear' / 'Good. This is all very useful information and will help us to help you' / 'O.K. We've nearly finished now'.

Chest
'Fine' / 'You have been very helpful so far' / 'You have been very helpful and I've nearly finished' / 'I understand' / 'O.K.' / 'I see' / 'Thank you'.

Rating the Encouraging Phrases

To get an idea about the strength of the E.P.s, four researchers were asked to rate them according to the amount of encouragement or strength of reinforcement which they judged them to contain. These ratings, in ordinal scale, are shown in Table 1. The 26 ratings of each of the five persons correspond to the 26 different Encouraging Phrases which were ordered as they occur in the four programs above. The Kendall coefficient of concordance showed a strong
First Person

Second Person

Third Person

Fourth Person

Table 1: Ratings of Encouraging Phrases (E.P.s)
agreement among the group ($X^2 = 71.13, \ p(0.001)$), with every pair of the group showing an agreement significant at the 0.001 level (Spearman rank correlation coefficient).

Based on the above ratings the graphical representations of the strength of E.P. = f (Question number) were drawn for each of the five programs (SEX was considered as two programs here, each corresponding to each sex), as shown in Figures 6, 7, 8, 9 and 10. Yet, preliminary inspection of those figures did not reveal any common pattern and it seemed as if randomness was the most probable case. (Statistical analysis on this matter is presented in section 2.5.1.)

All the E.P.s appear from their content to be independent of the meaning of the particular question to which they are attached. However, a small group of E.P.s (6.8%), e.g. 'Congratulations', 'Splendid', should follow an answer showing a personal success or a social advantage of the user. Similarly another small group of E.P.s (1.7%), e.g. 'I am very sorry to hear that', should follow an answer showing a social disadvantage. Finally, a third small group of E.P.s (6.8%), e.g. 'You have been very helpful and I have nearly finished', should be used after most, e.g. 80%, of the questions have been asked. The rest 85% of the E.P.s seem from initial inspection to be completely context-free and thus usable anywhere in a computer interview. (Statistical analysis on this matter is described in section 2.6.)

B) Phrases included in the question
This subject is discussed below in section 2.4.2 B).

C) Freely attached or detached phrases
These phrases just precede a question to give the impression of a natural dialogue, being freely attached to or detached from it.
The following phrases of this kind were found in each of the four programs (the right upper index indicating their frequency).

**Pregnancy**

'First of all' / 'For instance'.

**Gastric Ulcer**

'Well first of all' / 'Tell me' / 'First of all' / 'One final question which I would like you to think carefully about' / 'Could you tell me' / 'Now I would like you to think carefully about your answer to the next question'.

**Sex**

'Now first of all' / '(First) let me ask you' / 'Now the next question' / 'Now finally' / 'Tell me' / 'Next' / 'Now I want to ask you some more questions (of this kind)' / 'Firstly tell me' / 'Now a few more questions' / 'Now here's a rather important question' / 'Now' / 'Now I want to ask you' / 'Let me ask you just one or two more questions' / 'Firstly' / 'Here is another question' / 'This may be a difficult question but' / 'Then, finally a very difficult question but do your best' / 'Firstly let me ask you' / 'Now finally' / 'First' / 'Incidentally' / 'Well now for some more (questions). Again do your best to answer these truthfully'.

**Chest**

'By the way' / 'Tell me' / 'I'd like you to think carefully about this now' / 'First of all' / 'Now' / 'Could you tell me' / 'Now this is a very important question'.

The phrases of the type 'firstly' or 'first of all' were found following a phrase, which informs the user about the subject on which the following questions (minimum one) will be concentrated. These phrases can only be used preceding the first of a series of questions, which will attempt to elicit information from the user on a unique subject of the interview (e.g. smoking habits, in an interview about chest).

Phrases with the word 'finally' were found preceding the last of a series of questions about a unique subject of the interview, or preceding the last question of the interview, and they can be used only at these places.

The rest of the phrases could be used at any point of the interview preceding a question.
D) The rest of the pre-phrases

These pre-phrases are used mainly for four reasons:

a) To inform the user about the kind of information and the sort of questions which the system will attempt during the next few questions (minimum one) e.g. 'Now a question about catarrh'.

b) To give the user directions or ask for his co-operation, e.g. 'Please answer 'YES' or 'NO'. If you don't understand answer 'NO'. '

c) To explain why a particular question is asked, e.g. 'The reason I ask you that is because dislike comes after something has happened and I want to try to find out what has caused this dislike'.

d) To explain a term or a situation concerning the following question(s), e.g. 'By your family I mean your own parents, grandparents, brothers and sisters. (The family in this case does not include your other children, if you have any.)'

The majority of the phrases were of category a). For each program the following phrases were detected per category:

Pregnancy: a) 21, b) 3, c) 2, d) 1.

Gastric Ulcer: a) 7, b) 1, c) 0, d) 2.

Sex: a) 7, b) 2, c) 2, d) 7.

Chest: a) 21, b) 4, c) 0, d) 1.

From the above 4 categories only the category b) has context-free phrases, which could be used anywhere in the interviewing program. Each of the 4 programs contains the following phrases of category b).

Pregnancy
'Please answer 'YES' or 'NO'. If you are uncertain answer 'NO' because I am sure you will know if you had any of the illnesses' / 'Please bear with me' / 'Please answer 'YES' or 'NO'. If you don't know answer 'NO'.'
Gastric Ulcer
'Now I would like you to think carefully before you answer the next question'.

Sex
'You may find some of the questions a bit embarrassing, but nevertheless do be sure to answer them truthfully. If at any stage you do feel embarrassed, just sit back and remember that I am a computer' / 'Try to answer the questions with a 'YES' or 'NO' if you can'.

Chest
'Before you push one of the buttons I'd just like to remind you that if you don't quite understand what I mean by the questions, then press the '?' button' / 'I want you to type 'YES' or 'NO'. Please think carefully about this however before you push one of the buttons' / 'Please think carefully about this and tell me' / 'Please be absolutely honest about this and if you need time to think before you answer don't be afraid to do so, and read the questions as many times as you want to'.

Among the phrases preceding the question, as we have seen, the majority of E.P.s seem independent of the meaning of the questions, and so appears to be the case for the freely attached or detached phrases. Therefore the above phrases could be used easily by the proposed system.

The next step then would be to examine the rest of the L.S.E., and to find out to what extent they are related to the meaning of the questions.

2.4.2 Questions differing from S.S.Q.s

This section reports on the several ways of asking a question as observed in the four medical interviewing programs which were analysed.

These several types of questions were used in order to increase the variability of the computer interview, and were intended to have a positive effect on the user's acceptance of it.
The possibility of transforming S.S.Q.s and vice-versa is also examined and transformational rules to enable this change are stated, when possible.

The transformational rules are discussed in this section, because it would be possible to incorporate them in the proposed computer system, should they be simple enough for this purpose.

A) If-Then Questions

The questions of this type and their transformations to S.S.Q. for each program are as follows: (the number is the serial number of the question in the particular program sequence).

Pregnancy
23) 'If you have taken any other medicines or drugs at all please press the 'YES' button'. § 'Have you taken any other medicines or drugs at all?'
98) 'If this took place more than 15 years ago press the 'YES' button'. § 'Did this take place more than 15 years ago?'

Gastric Ulcer
18) 'If you do (find the interview interesting) push the 'YES' button otherwise push the 'NO' button'. § 'Do you (find the interview interesting?)'.
23) 'If you had to describe this liquid or fluid would you say it was clear and perhaps tasted salty?' § 'Would you say that this liquid or fluid was clear and perhaps tasted salty?' § (Transformation of questions with embodied pre-phrases.) 'Was this liquid or fluid clear and perhaps tasted salty?'.

Sex
70) 'If you have a husband or boyfriend at the moment do you love him?'. (No transformation necessary.)
79) 'If he cannot come it would be helpful if you come first, will you do this?'. § 'If he cannot come, would you come first to help?'.

Chest
25) 'If you use some other method, press the 'YES' button, but if not push the 'NO' button'. § 'Do you use some other method?'. 
40) 'If any other kind of weather, which I haven't mentioned, affects your chest, push 'YES'. If no push 'NO'. 'Does any other kind of weather, which I haven't mentioned, affect your chest?'.

60) 'If you inhale the smoke slightly press the 'YES' button. 'Do you inhale the smoke slightly?'.

91) 'If you smoke cigarettes and roll your own and a pipe and large cigars and small cigars every day, push 'YES'. If only on some days press 'NO'. 'Do you smoke cigarettes and roll your own and a pipe and large cigars and small cigars every day?'.

The following deductions and comments seem appropriate.

When transforming a S.S.Q. to an 'If-Then' question the word 'then' is usually omitted, as it is implied in the context.

These 'if-then' questions can be further classified into two groups:

1) the part after 'then' is a direction (e.g. 'Press the 'YES' button', or 'Press the 'YES' button, but if not push the 'NO' button').

2) The part of the question following 'then' is not a direction. The majority of these 'if-then' questions belong to the 1st group (seven out of ten).

The questions of group 1 can easily be transformed to S.S.Q. by a transformation of the surface structure of the sentence. This transformation consists in discarding the last part of the sentence including 'then' and manipulating the first part as follows. The word 'if' is omitted and the interrogative form is used, by using the proper auxiliary in the same form of the main verb as in the 'if-then' question.

The questions of group 2 cannot in general be transformed to S.S.Q. with a transformation of their surface structure elements. Sometimes they consist of the S.S.Q. itself (e.g. Sex 70 above); sometimes they can be better transformed to S.S.Q. through a transformation on the deep structure (e.g. Sex 79 above). Sometimes they can be transformed to S.S.Q.s taking into account elements of the surface structure of the sentence, but again not without looking at their meaning. Consequently it is not possible to formulate a general rule, based only on the position of certain syntactic and grammatical terms, which will transform questions of group 2 to S.S.Q. and vice-versa.
8) Questions with included phrases

The included phrases are context-free as far as the meaning of the underlying S.S.Q. is concerned. They are attached to the S.S.Q. resulting in a change, which would be represented by a simple transformation on the elements of the surface structure of the beginning of the S.S.Q.

For a question with an included pre-phrase to be transformed to a S.S.Q. the pre-phrase is detached, and the interrogative form replaces the affirmative one. (The proper auxiliary is used in the same tense of the verb as in the affirmative form, followed by the subject and the main verb.) If a time indicator exists within the pre-phrase (often, sometimes, etc.), then it is used between the subject and the main verb in the corresponding S.S.Q. The reverse manipulation transfers the S.S.Q. to a question with an included pre-phrase.

Two types of these pre-phrases were found: a) Ending with if or whether. b) Ending with that (which can be omitted). The following phrases were observed for each type:

a) 'Would you please tell me if'
   'I would like to know if'
   'I wonder if'

b) 'Would you say (that)'
   '(Now) do you find that'
   'Do you sometimes find (that)'
   'Have you ever been told that'

The phrases of each program together with their transformation to a S.S.Q. are as follows.

Pregnancy
a) 'Would you please tell me if you have every had ...' ⇋
   'Have you ever had ...' / 'I would like to know if your confinement took place ...' ⇋ 'Did your confinement take place ...'
Gastric Ulcer

a) 'Would you say that you were eating ...?' \( \Rightarrow \) 'Were you eating ...'? 'Now do you find that it gives you ...'? 'Does it give you ...'? 'Do you sometimes find that your mouth unexpectedly fills up ...'? 'Does your mouth unexpectedly fill up ...?'

Sex

a) 'I wonder if you both will come ...'? 'Will you both come ...'? 'Now I want to know if you like ...'? 'Do you like ...'? 'I wonder if you had realised ...'? 'Had you realised ...'? b) 'Do you find that you get angry ...'? 'Do you get angry ...'? 

Chest

a) 'Would you say that, when you get a cold, it goes down ...'? 'When you get a cold, does it go down ...'? 'Would you say it only takes ...'? 'Does it only take ...'? 'Would you say that you bring up ...'? 'Do you bring up ...'? 'Do you find that dry weather actually seems ...'? 'Does dry weather actually seem ...'? 'Would you say that you had this ...'? 'Did you have this ...'? 'Have you ever been told that you have ...'? 'Have you ever had ...'? 'Would you say that you started ...'? 'Did you start ...'? 'Would you say you were ...'? 'Were you ...'? 'Would you say it was ...'? 'Was it ...'? 'Would you say that you smoked ...'? 'Did you smoke ...'? 

Two particular features may be noted from these examples.
If a time phrase follows the pre-phrase then it precedes the S.S.Q. also, while the transformation takes place after this time phrase.
The phrase 'Have you ever been told that' usually causes the main verb of the S.S.Q. to be in the Present Perfect tense.

C) Descriptive questions

The questions of this kind for every program, and the way they are transformed to S.S.Q., are as follows.

Pregnancy

14) 'Sometimes a period can be heavier than usual and sometimes can be lighter. Was your last period unusual in this way? ' 'Was your last period heavier or lighter than usual?' 

Gastric Ulcer

None.
Sex

1) 'I gather that you have come for help because of some problem with your sexuality or your sex life. Is this correct?' ⇔ 'Have you come for help because of some problem with your sexuality or your sex life?'

4) 'Now another thing which I'd like to tell you is that in many cases sexual problems can be treated relatively easily. I wonder if you had realised this - had you?' ⇔ 'Had you realised that in many cases sexual problems can be treated relatively easily?'

5) 'It will be of great assistance to us in our efforts to help you if you answer these questions to the best of your ability. Please will you try to do this?' ⇔ 'Would you please try to answer these questions to the best of your ability (which would be of great assistance to us in our efforts to help you)?'

Chest

33) 'Some people find that the weather affects their chests. Do you find this in your own case?' ⇔ 'Does weather affect your chest?'

These phrases cannot in general be transformed to S.S.Q.s taking into account only the positions of elements of the surface structure of the question. The phrases Sex 1) and Chest 33) depend slightly on the meaning of the question. It is felt that descriptions of this type could fit with S.S.Q.s when the main verb is 'suffer from' or 'have', e.g. 'Sometimes people suffer from 'something'. Do you have the same problem?' ⇔ 'Do you suffer from 'something'?'

2.4.3 Phrases following the question

The following section describes the several phrases, which were used after the main question in the four interviewing programs, which were analysed.

Moreover the relation of these phrases to the meaning of the main questions, and consequently the possibility of their use in the proposed computer system, is discussed.
A) Post-phrases explaining or analysing a part of the question
These phrases are context-sensitive depending on the meaning of a part of the question. In most cases this part is a medical term. They were found as follows: Five in Pregnancy (numbers 48, 49, 59, 60, 67), One in Gastric Ulcer (number 15), and one in Sex (number 56).

B) Post-phrases directing the user
The following phrases of this kind were found in the four programs.

Pregnancy
95) 'Press the 'YES' if it was either of these hospitals'.

Gastric Ulcer
3) 'Push 'YES' or 'NO'.'
5) 'Please push 'YES' or 'NO'.'
9) 'Push 'YES' if you know where the navel is and 'NO' if you don't know'.
10) 'Think carefully and push 'YES' or 'NO'.'
12) 'Push 'YES' if it does and 'NO' if it doesn't'.
13) 'Press 'YES' or 'NO'.'
14) 'Push one of the buttons'.
19) 'Push 'YES' if you have and 'NO' if not'.
21) 'Press one button'.
22) 'Press 'YES' or 'NO' please'.
24) 'Press 'YES' if it does, and 'NO' if not'.
26) 'YES' or 'NO'.'
27) 'If you do push the 'YES' button, if not push the 'NO' button'.
28) 'Press 'YES' if you have and 'NO' if not'.

Sex
1) 'Push one of the buttons'.
2) 'Push 'YES' or 'NO'.'
15) 'Push 'YES' or 'NO'.'
53) 'Push 'YES' or 'NO'.'

Chest
1) 'Press one of the buttons now'.
3) 'Push one of the buttons now'.
4) 'Push one of the buttons'.
12) 'If you're not quite sure what I mean push the '? button'.
19) 'Answer 'YES' or 'NO' (if you don't normally need to clear your chest at all, you should answer 'YES')'.
56) 'Push 'YES' if you have, 'NO' if you haven't and '?' if you are not sure'.

57) 'If you're not sure whether to describe yourself as a smoker or not, press the '?' button'.
60) 'O.K., go ahead and press one of the buttons when you've decided'.
104) 'Push 'YES' if you have and 'NO' if you haven't'.
109) 'Press 'YES' if you have and 'NO' if not'.

In some of these phrases the '?' is given a meaning other than the 'I don't understand' one. This situation seems not be helpful for the automation of medical interviews, which is one of the aims of the proposed system for which this analysis has been attempted. For this reason this special rare case has not been taken into account. The post-phrases directing the user can be further classified into two groups: 1) Post-phrases containing directions only, 2) Post-phrases referring back to the question in order to explain under which condition the subsequent instruction is valid. The group 1) includes 17 out of the 29 post-phrases directing the user. The phrases are context-free and can be used anywhere in the program. The group 2) includes the remaining 12 of the 29 phrases shown above. The phrases of this group are context-sensitive.

However there is one case when the phrases of group 2) can be formed without looking at the meaning of the question, but only by identifying certain elements of its surface structure, namely the subject and the auxiliary verb of the question. This is when the reference to the question consists only of the subject and the auxiliary is preceded by the word 'if' e.g. Chest 104 (above). The phrases of this case were found to be 8 out of the 12 phrases of group 2).

Therefore 25 out of the 29 phrases directing the user (86%) could be used during a medical interview, without any knowledge of the meaning of its questions. The total percentage of the phrases following the question, which are independent of the meaning of the main question, and therefore could be easily be used by the proposed computer system, is about 70%.
From what has been said so far for the L.S.E., it can be concluded that the majority of them (about 70%) seem, from preliminary inspection, to be independent of the meaning of the underlying S.S.Q.s. Therefore these 70% of the L.S.E. could be used in the proposed computer system.

Yet before testing the usefulness or not of these phrases in a computer interview two more operations would be necessary:
1. to find out the patterns, which the L.S.E. follow among the questions,
2. to test further whether the seemingly context-free L.S.E. bear any relation to the meaning of the question, probably by using statistical analysis.

2.5 Patterns of L.S.E.

Section 2.5 describes a part of the research which aimed at finding what patterns the L.S.E. follow among the questions. These patterns could then be used by an algorithm to convert a questionnaire to an acceptable interviewing program.

2.5.1 Patterns of Encouraging Phrases (E.P.s)

The following analysis consists of an effort to find out if the E.P.s, distributed among the questions of the 5 examined programs as shown in figures 6-10, followed any common patterns or not, and if so, to define these patterns. This was attempted by generating and testing for significance statistical hypotheses concerning the data of figures 6-10.

As no theory exists, which would guide the generation of hypotheses relevant to this kind of data, the hypotheses were based on
1) the observation of figures 6-10 for common patterns and tendencies, and
2) the estimation of what might be more likely, for example, defining the E.P.s of level < 6 as 'mild reinforcements' and those ≥ 6th level among 11 possible levels as 'strong ones', one would expect that there is a higher probability for a 'mild reinforcement' to occur than a 'strong one'.

Working this way three hypotheses were generated and then tested:

1) The E.P.s are randomly distributed among the questions.
2) There is a higher probability for a 'mild reinforcement' to occur than a 'strong one'.
3) The 'mild reinforcements' and the 'strong reinforcements' differ in the proportion they fall into the two halves of each of figures 6-10, with increasing proportion in the second half.

The first hypothesis was tested by a runs test (after carefully examining the diagram Strength of E.P. = f (Question number), and observing no obvious deviations from randomness, and was found to be acceptable for all five programs apart from SEX (females), which just fell short of the relevant requirements. Therefore it seems that the E.P.s do not follow any common pattern among the questions.

The second hypothesis was tested by a binomial test, and was found to be valid for two among the five questionnaires (SEX (males), and CHEST), while the reverse hypothesis of equal probability for a 'mild' and a 'strong reinforcement' to occur (null hypothesis), cannot be rejected for the other three programs.

The third hypothesis was tested by a Fisher test and a Rank correlation test. The result was that the reverse hypothesis, i.e. that the 'mild' and 'strong reinforcements' do not differ in the proportion they fall into the two halves of each diagram (null hypothesis), cannot be rejected for any of the five programs.

The level of significance for testing all the above hypotheses was 0.05.

As a conclusion of the above study it could be said that, though the data are not the most suitable for statistical treatment because of their small number, it seems that the E.P.s follow no common patterns among the questions and that randomness is the most probable case.
2.5.2 Patterns of the rest of L.S.E. (apart from E.P.s)

The E.P.s (Encouragement) seem to follow a rather random pattern among the questions, as the previous analysis has indicated.

As far as the rest of the L.S.E. are concerned, these were incorporated into a second variable called 'Chattiness' (Chatty Phrases). This variable consisted mainly of freely attached or detached phrases, questions differing from S.S.Q.s and phrases directing the user (mainly reminding him/her how to communicate with the system).

To study the patterns, which the Chatty Phrases (C.P.s) follow among the questions, the questions of the five programs (SEX was again considered as two programs for this analysis, one for Males and the other for Females) were classified as chatty ones and non-chatty ones. For this classification the questions which contained chatty phrases were characterised as chatty questions, while the rest, that is the S.S.Q.s and the S.S.Q.s with E.P.s, were classified as not chatty ones. Figures 11-15 of the Appendix show how the chatty phrases were distributed among the questions.

Then the hypothesis that the Chatty Phrases (chatty questions) were randomly distributed among the questions was tested by a runs test (after carefully examining first the diagrams C.P.s = f (Question number) and observing no obvious deviations from randomness). The result was that this hypothesis is acceptable for all five programs at the 0.05 level of significance.

Therefore it seems that the C.P.s, like the E.P.s, do not follow any common patterns among the questions and that, again, randomness is the most probable case.
2.6 Correlation between L.S.E. and the Meaning of the Questions

Preliminary analysis of the four history-taking interviewing programs indicated that about 70% of the L.S.E. involved in these programs appeared not to be context-sensitive and thus could be used in the computer software system envisaged to convert questionnaires into interview programs. In particular most of the E.P.s (about 85%) appeared from their content to be context-free, while the context-free among the chatty phrases were either freely-attached or detached phrases, questions differing from S.S.Q.s or phrases directing the user. Moreover for most of the questions differing from S.S.Q.s (apart from the Descriptive ones), simple transformational rules could be derived enabling their transformation to S.S.Q.s and vice-versa (as was stated in section 2.4.2).

To investigate further whether these seemingly context-free L.S.E. bear any relation to the meaning of the questions, the following procedure was adopted.

First, four variables were strictly defined. These variables were thought to be the most important ones in the interview situation by a panel of four researchers and the author. The four variables and their definition are as follows:
1) Sensitivity: the extent to which a question deals with personal matters and is likely to generate an emotional response in the interviewee.
2) Confidentiality: the degree to which the interviewee might want his answer to a question kept secret.
3) Subjectivity: the degree to which a question deals with feelings, opinions, imaginary events, etc., rather than with objective (or factual) data such as age.
4) Dissimilarity: the degree to which the subject matter of a question is different from that of the previous question.
Then the questions of the four medical interviewing programs (apart from PREGNANCY) were rated by the panel of four researchers (males) on each of the four variables. Tables 2 and 3 show these ratings for programs SEX (males) and SEX (females) correspondingly.

The agreement among the four persons of the panel was tested by the Kendall coefficient of concordance. The test showed significant results i.e. agreement among the panel, at the 0.02 level of significance of all the four variables for the programs SEX (males) and SEX (females). For the other two programs (GASTRIC ULCER and CHEST) the small degree of variation did not permit a reliable test of agreement among the panel members; however, it was obvious by observing the raw data that there was a good agreement among the four researchers.

Tests were then made for the program SEX (Males) and SEX (Females) using the Kendall correlation of concordance, to find the correlation between the degree of encouragement in the Encouraging Phrases and each of the four variables, and the correlation of the degree of verbosity of each question with each variable. For these tests each E.P. was represented by the median of the ratings of the four members of the panel. Similarly the median was taken of the ratings on each variable for all the four variables and all the questions of each of the four programs. The total number of words of each question was taken as an indication of its verbosity. The result of those tests was: no correlation for both groups of data at any acceptable level of significance.
Table 2: Ratings of the 4 variables for program SEX (Males)

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Sensitivity Persons</th>
<th>Confidentiality Persons</th>
<th>Subjectivity Persons</th>
<th>Dissimilarity Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>1</td>
<td>3 1 1 2</td>
<td>2 2 1 1</td>
<td>3 4 2 3</td>
<td>2 2 4 1</td>
</tr>
<tr>
<td>2</td>
<td>3 1 1 2</td>
<td>2 2 1 1</td>
<td>3 4 2 3</td>
<td>1 1 2 1</td>
</tr>
<tr>
<td>3</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>2 4 2 2</td>
<td>3 4 2 4</td>
</tr>
<tr>
<td>4</td>
<td>2 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>5 3 3 4</td>
</tr>
<tr>
<td>5</td>
<td>4 4 2 3</td>
<td>3 4 2 3</td>
<td>4 4 2 4</td>
<td>5 3 4 4</td>
</tr>
<tr>
<td>6</td>
<td>4 4 2 3</td>
<td>3 4 2 3</td>
<td>4 4 2 4</td>
<td>2 3 1 1</td>
</tr>
<tr>
<td>7</td>
<td>4 5 3 4</td>
<td>5 5 3 4</td>
<td>3 4 2 4</td>
<td>4 2 2 2</td>
</tr>
<tr>
<td>8</td>
<td>4 4 2 5</td>
<td>3 5 2 5</td>
<td>4 5 2 4</td>
<td>5 2 2 2</td>
</tr>
<tr>
<td>9</td>
<td>4 4 2 4</td>
<td>4 5 2 3</td>
<td>4 4 2 4</td>
<td>3 2 2 2</td>
</tr>
<tr>
<td>10</td>
<td>4 4 3 5</td>
<td>4 5 3 4</td>
<td>4 5 2 4</td>
<td>2 1 2 2</td>
</tr>
<tr>
<td>11</td>
<td>4 5 3 5</td>
<td>4 5 3 5</td>
<td>4 4 2 5</td>
<td>3 1 2 2</td>
</tr>
<tr>
<td>12</td>
<td>5 4 2 5</td>
<td>5 3 2 5</td>
<td>4 4 2 4</td>
<td>5 3 2 3</td>
</tr>
<tr>
<td>13</td>
<td>3 3 2 2</td>
<td>2 3 2 2</td>
<td>4 5 2 3</td>
<td>4 3 2 4</td>
</tr>
<tr>
<td>14</td>
<td>5 3 4 5</td>
<td>4 4 3 4</td>
<td>3 4 2 4</td>
<td>5 4 3 4</td>
</tr>
<tr>
<td>15</td>
<td>4 3 4 4</td>
<td>4 4 3 3</td>
<td>4 4 2 4</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>16</td>
<td>4 4 2 5</td>
<td>5 4 2 5</td>
<td>1 2 2 2</td>
<td>2 2 1 3</td>
</tr>
<tr>
<td>17</td>
<td>5 4 2 5</td>
<td>5 4 2 5</td>
<td>1 2 2 2</td>
<td>1 1 1 2</td>
</tr>
<tr>
<td>18</td>
<td>4 2 2 3</td>
<td>3 2 1 2</td>
<td>3 2 2 2</td>
<td>3 2 1 2</td>
</tr>
<tr>
<td>19</td>
<td>4 2 2 2</td>
<td>3 2 1 2</td>
<td>3 2 2 2</td>
<td>2 1 1 1</td>
</tr>
<tr>
<td>20</td>
<td>4 4 3 3</td>
<td>5 5 2 3</td>
<td>4 4 2 2</td>
<td>3 2 1 2</td>
</tr>
<tr>
<td>21</td>
<td>5 4 3 4</td>
<td>4 5 2 4</td>
<td>2 3 2 2</td>
<td>5 3 4 3</td>
</tr>
<tr>
<td>22</td>
<td>4 4 3 4</td>
<td>3 4 3 4</td>
<td>2 3 2 2</td>
<td>2 1 1 3</td>
</tr>
<tr>
<td>23</td>
<td>4 3 3 5</td>
<td>4 4 3 5</td>
<td>3 2 2 3</td>
<td>2 2 1 2</td>
</tr>
<tr>
<td>24</td>
<td>4 3 3 3</td>
<td>4 4 3 3</td>
<td>3 2 2 2</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>25</td>
<td>5 3 3 3</td>
<td>3 3 3 3</td>
<td>5 4 2 2</td>
<td>5 2 4 3</td>
</tr>
<tr>
<td>26</td>
<td>5 3 4 3</td>
<td>4 4 3 2</td>
<td>4 4 2 3</td>
<td>2 1 2 2</td>
</tr>
<tr>
<td>27</td>
<td>5 4 4 4</td>
<td>4 5 3 5</td>
<td>2 2 2 2</td>
<td>3 2 3 2</td>
</tr>
<tr>
<td>28</td>
<td>4 4 4 4</td>
<td>5 4 3 4</td>
<td>3 2 2 2</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>29</td>
<td>4 4 4 5</td>
<td>4 5 3 5</td>
<td>3 2 2 2</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>30</td>
<td>4 2 3 2</td>
<td>2 2 3 2</td>
<td>5 1 2 4</td>
<td>5 4 4 3</td>
</tr>
<tr>
<td>31</td>
<td>5 1 3 3</td>
<td>2 3 3 3</td>
<td>5 4 2 5</td>
<td>2 1 2 1</td>
</tr>
</tbody>
</table>
Table 3: Ratings of the 4 variables for program SEX (Females)

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Sensitivity Persons</th>
<th>Confidentiality Persons</th>
<th>Subjectivity Persons</th>
<th>Dissimilarity Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 2</td>
<td>1 2 1 1</td>
<td>5 4 2 2</td>
<td>2 5 4 1</td>
</tr>
<tr>
<td>2</td>
<td>1 1 1 2</td>
<td>1 2 1 1</td>
<td>4 4 2 2</td>
<td>5 1 2 1</td>
</tr>
<tr>
<td>3</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>5 4 2 1</td>
<td>3 4 2 3</td>
</tr>
<tr>
<td>4</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>3 3 3 3</td>
</tr>
<tr>
<td>5</td>
<td>4 3 2 3</td>
<td>3 4 2 3</td>
<td>5 5 2 4</td>
<td>4 4 4 2</td>
</tr>
<tr>
<td>6</td>
<td>3 3 3 3</td>
<td>3 4 3 4</td>
<td>5 5 2 4</td>
<td>2 1 2 1</td>
</tr>
<tr>
<td>7</td>
<td>3 4 3 4</td>
<td>3 4 3 4</td>
<td>4 5 2 4</td>
<td>3 2 3 2</td>
</tr>
<tr>
<td>8</td>
<td>4 5 4 5</td>
<td>4 5 4 5</td>
<td>4 5 2 5</td>
<td>3 1 4 1</td>
</tr>
<tr>
<td>9</td>
<td>4 4 3 3</td>
<td>4 5 3 5</td>
<td>3 5 2 4</td>
<td>3 2 3 1</td>
</tr>
<tr>
<td>10</td>
<td>4 3 3 5</td>
<td>3 4 3 5</td>
<td>4 5 2 5</td>
<td>2 1 3 2</td>
</tr>
<tr>
<td>11</td>
<td>3 4 3 4</td>
<td>2 5 3 4</td>
<td>4 5 2 4</td>
<td>4 2 2 1</td>
</tr>
<tr>
<td>12</td>
<td>2 2 3 3</td>
<td>1 1 3 3</td>
<td>3 5 2 3</td>
<td>3 3 2 3</td>
</tr>
<tr>
<td>13</td>
<td>3 4 2 3</td>
<td>3 5 2 3</td>
<td>4 4 2 4</td>
<td>5 3 4 2</td>
</tr>
<tr>
<td>14</td>
<td>4 4 3 4</td>
<td>4 4 3 4</td>
<td>4 4 2 5</td>
<td>2 1 2 2</td>
</tr>
<tr>
<td>15</td>
<td>5 5 4 5</td>
<td>4 5 4 5</td>
<td>4 4 2 5</td>
<td>3 1 3 1</td>
</tr>
<tr>
<td>16</td>
<td>4 5 4 5</td>
<td>5 5 3 5</td>
<td>2 3 2 5</td>
<td>3 2 2 1</td>
</tr>
<tr>
<td>17</td>
<td>4 5 4 5</td>
<td>4 5 3 5</td>
<td>3 2 2 5</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>18</td>
<td>4 4 4 5</td>
<td>4 5 3 5</td>
<td>4 4 2 5</td>
<td>2 1 1 1</td>
</tr>
<tr>
<td>19</td>
<td>4 4 4 5</td>
<td>4 5 3 5</td>
<td>4 4 2 5</td>
<td>3 1 1 1</td>
</tr>
<tr>
<td>20</td>
<td>3 3 3 5</td>
<td>2 4 3 5</td>
<td>3 3 2 5</td>
<td>4 2 4 2</td>
</tr>
<tr>
<td>21</td>
<td>4 3 3 5</td>
<td>5 4 3 4</td>
<td>4 3 2 4</td>
<td>2 2 2 2</td>
</tr>
<tr>
<td>22</td>
<td>4 3 3 2</td>
<td>2 4 2 2</td>
<td>5 4 2 4</td>
<td>5 4 4 4</td>
</tr>
<tr>
<td>23</td>
<td>4 3 3 3</td>
<td>2 4 2 3</td>
<td>5 5 2 4</td>
<td>1 1 2 1</td>
</tr>
</tbody>
</table>
Because of the small degree of variation of some of the programs (GASTRIC ULCER and CHEST) in the four variables, and to enable a detailed comparison of the patterns in the groups of data, the following graphical representations were drawn for each of the four programs.

(1) \( E.P. = f (Q.N.) \)
(2) \( S_j = f (Q.N.) \)
(3) \( T.N.W. = f (Q.N.) \)
(4) \( E.P. = f (S_j) \)
(5) \( T.N.W. = f (S_j) \)

where:
- \( E.P. \) = Strength of Encouragement of an Encouraging Phrase
- \( f \) = function of
- \( Q.N. \) = question number
- \( S \) = symbol representing one of the four variables
- \( j = 1, 2, 3, 4 \) index referring to the four variables
- \( T.N.W. \) = total number of words

The diagrams represented by equations (4) and (5) were not drawn for program CHEST. The reason was that because of the big number of questions of this program (144) and the lack of variation of the data, the diagrams (4) and (5) would not result in a very clear picture. Therefore these diagrams were not used for program CHEST in the following comparisons. Examples of the diagrams described by equations (2), (3), (4) and (5) are given for program GASTRIC ULCER in figures 16-28. The rating scores of the four researchers have been ranked in order to remove their individual bias. This ranking would ensure that the scores of the four researchers are expressed on the same basis, even if one of them consistently rated the questions towards the low (or high) end of the scale.

Comparing (1) and (2) for every program and all the variables showed no discernable pattern relating \( E.P.s \) and the four variables and this indication was further reinforced by observing (4). Exactly the same results, that is no discernable pattern, were shown for Chattiness and each of the four variables either by comparing (3) and (2) or by observing (5).
Thus for these four interview programs, their authors do not seem to use the placing of Encouraging Phrases or the length of the questions to indicate or relate to the meaning of the questions. If this is generally true, then it will be much easier to plan a computer program to generate interviews, which will be equivalent to such interviews written by human programmers.

2.7 Discussion on the Analysis

The analysis described so far was mainly undertaken to collect useful information about the linguistic content of acceptable computer interview questionnaires. This information could then be incorporated in a computer system to carry out 'natural' and 'friendly' interviews. To fulfil the above aim four medical history-taking interview programs were analysed from grammatical, syntactic and semantic points of view. These programs were chosen because their authors deliberately emphasised the style of the questions, in order to increase their acceptability to the interviewees. Moreover the four programs analysed were successfully tested on actual patients (as described in section 1.1.2). This task was attempted without entering the full complexity of the semantics of the language.

Consequently during the general analysis the Linguistic Style Elements (L.S.E.) used in the interview questionnaires were identified and classified, and an investigation was attempted on the patterns which the L.S.E. follow among the questions, and on whether the L.S.E. are related to the meaning of the questions. The problem arising when a patient interviewee does not understand a question was also considered as a secondary part of the analysis. The analysis of the four computer interviewing programs has revealed a number of general features which are discussed below.
The flow charts of the programs, as shown in figures 2-5, present a central trunk and closed side branches returning to some point of the main trunk. The different complexity of the flow charts of the programs reflects the difference of familiarity with computer programs of the several persons who wrote them.

There is a rather long introduction, with the aim that the patients become accustomed to the computer and feel calm and ready to answer the main questions which follow.

The 'friendly' and 'natural' way of carrying out the interview is established through L.S.E. The L.S.E. consist of several phrases and questions differing from S.S.Q.s.

The L.S.E. were divided into three parts: (1) the phrases preceding the question, (2) the question itself, if different from S.S.Q., and (3) the phrases following the question. These three categories were then divided into sub-categories.

From the several groups of pre-phrases, special attention was paid to the Encouraging Phrases, aiming to improve the patient's positive attitude towards the interview. These phrases were classified in several levels of strength of 'reinforcement'. Statistical tests and analytical diagrams have indicated that the E.P.s follow no common or discernable pattern among the questions of the four examined programs; therefore they appear to be placed at random in the sequence of a program. Another group of pre-phrases, the group of freely attached or detached phrases, consists of context-free phrases which could be used in the proposed system in the same way as most of the Encouraging Phrases. The group of phrases included in the questions consists also of context-free phrases. From the rest of the pre-phrases only the ones directing the user are suitable for the proposed software system, the rest being context-sensitive.
With regard to part (2) of the L.S.E., the question itself, the group of 'If-Then' questions consists of questions which can be transformed to S.S.Q.s and vice-versa through a simple transformation of their surface structure elements. This is also the case for the group of questions with included phrases. In this case the transformation is even simpler, applying to the surface structure elements of the beginning of the sentence. Therefore both these groups are suitable for the proposed system. The last group of part (2), that is the group of descriptive questions, can generally be transformed to S.S.Q.s and vice-versa only in special cases.

Part (3), the category of phrases following the questions, consists of two groups: (a) post-phrases analysing a part of the question, and (b) post-phrases directing the user. The first group includes context-sensitive elements, while the second one includes mainly context-free elements suitable for the proposed system.

The variable Encouragement (E.P.s) seems to follow a random pattern among the questions, as already analysed and discussed. To study the pattern of the rest of the L.S.E., these were incorporated into a second variable the Chattiness (Chatty Phrases). Then a statistical runs test has indicated that the Chatty Phrases (C.P.s) are rather randomly distributed among the questions too.

To test the seemingly context-free L.S.E. further, if they imply any relation to the meaning of the questions, the questions of the four programs were rated on each of four scales, which were thought to be important in the interview situation. Yet, no significant correlation among the degree of Encouragement and the verbosity of the four questionnaire programs, and the four scales (SENSITIVITY, CONFIDENTIALITY, SUBJECTIVITY, DISSIMILARITY) was found.
For the problem resulting when a patient (interviewee) does not understand a question, two possible solutions were examined: first, the possibility of changing a phrase using standard forms of questions and a look-up table; second, the explanation of a part of the question, which probably causes the trouble. The first option was rejected in favour of the second on the grounds that the medical questions, though they present certain common points in their structure, do not consist of an organised group which would permit the existence of certain simple standard forms of questions. Therefore a stored dictionary of medical terms (technical terms for an interview of any kind) was suggested, to state the explanation of the 'difficult' terms involved in a question, should the interviewee not understand that particular question.
3. RATIONALE FROM THE ANALYSIS AND PLANS OF THE EXPERIMENTAL PROGRAMME

This chapter reviews the results of the analysis described in Chapter 2, shows how these results relate to the general research aims, and sets out the philosophy and plan for the main research programme.

3.1 Rationale from Chapter 2 regarding Research Aims

The analysis of Chapter 2 was undertaken to aid our understanding of the computer interview process and in particular of the linguistic contents of an acceptable interviewing program. The latter was the first aim of this research.

The analysis has revealed the L.S.E., i.e. phrases preceding or following the questions and various ways of asking the questions. The L.S.E. were used to increase the variability of the interaction, and make the computer interview more friendly and chatty, and therefore more acceptable to the interviewees. It has also revealed that the L.S.E. follow rather random patterns among the questions. Furthermore the analysis has shown that about 70% of the L.S.E. were independent of the meaning of the simple and straightforward questions.

The above findings encourage the possibility of the construction of a computer interviewing system, to carry out acceptable interviews, based on the findings of the analysis. The construction and evaluation of such a system consists of the second aim of this research.

Finally the adequacy of such a system based only on the context-free L.S.E. should be investigated, as the third aim of this research demands.
3.2 General Plan and Experimental Programme

However, the question which clearly comes to mind following the above analysis is whether the L.S.E. are useful at all, and if so, how. For example, do people prefer interviews which include more or fewer encouraging phrases? Again, if they receive a style of interview which they prefer, do they give no less accurate answers to the questions?

To answer questions of this kind several tests are necessary. The task of doing these tests seems complicated and multi-dimensional. For example, the L.S.E. should be tested for increase in acceptability to the users, for accuracy of the derived information, for probable changes in the users' reaction time, etc. The questionnaire of S.S.Q.s could be tested for acceptability against several versions of the same basic questions, but with varying degree of Encouragement and Chattiness, the corresponding phrases (E.P.s and C.P.s) following a random pattern as the analysis in Chapter 2 has indicated. The random patterns of E.P.s and C.P.s could also be tested against other more planned patterns (e.g. a periodic pattern or one with increasing reinforcement towards the end of the interview, where the interviewee's concentration is rather expected to decrease etc.).

It would also be interesting to conduct an experiment on the effect of L.S.E. on interview acceptability in a man-man interviewing situation and compare the results with those of the man-computer situation.

Therefore five experiments were conducted aiming to evaluate the effect of L.S.E. on people's acceptance of computer and human interviews.
These five experiments, which refer to measurements of people's attitude, are described in Chapters 4-8 correspondingly. Since this study focuses on computer interviewing, four out of the five experiments were related to computer interviews and only one used human interviews, the latter being discussed in Chapter 6. Chapter 9 consists of a discussion of the results of the experiments described in Chapters 4-8.

Finally it will be essential to carry out an experiment in order to assess the degree of reproducibility and accuracy of the information derived from the computer interviews, possibly by comparing it with the same information derived through human interviews. This experiment is described in Chapter 10, while Chapter 11 discusses the results of the experiment.

3.3 Methods

The six experiments, described in Chapters 4-11, had a lot in common in the methodology which they used in order to achieve their aims (which are discussed in sections 4.1, 5.1, 6.1, 7.1, 8.1 and 10.1 correspondingly). The subject was divided in several sections, e.g. subjects, variables, etc., for each of the six experiments as summarized below.

3.3.1 Material

The basic material used for all six experiments was a questionnaire about leisure time and people's hobbies. A more detailed description of this basic questionnaire is given in section 4.2.

3.3.2 Variables

For three of the six experiments (described in Chapters 4, 5 and 6) the amount of random E.P.s and C.P.s was used as the controlled variable, while for one experiment (described in Chapter 7) the pattern of E.P.s and C.P.s was used for this purpose.
The experiment described in Chapter 8 used the 'version of the questionnaire' (as presented to the interviewees by human interviewers and the computer) as its controlled variable, while the last experiment (described in Chapter 9) used the method of interviewing (i.e. human or computer interviews) as its controlled variable.

The design of four of the experiments was between- rather than within- subjects, while for the remaining two experiments (described in Chapters 8 and 10) the within-subjects design was used.

3.3.3 Subjects

The subjects of the experiments were either people from the general public (for two out of six, one with computer interviews and the other with human interviews), or students and young apprentices (for the other four experiments).

3.3.4 The Computer System

The computer system used during the experiments was capable of working in two stages: a) it could accept a questionnaire version and convert it to the interviewing program, and b) it could carry out the interview, keeping trace of the questions asked, the answers given, the reaction times, etc.

This subject is presented in more detail in section 4.6.

3.3.5 The Man-Computer and the Man-Man Interaction

This section describes the way the questions were presented to the subjects. Accordingly, during the computer interviews the computer presented the questions on a V.D.U. in natural English, while the interviewees were communicating with the system by pressing the three buttons of a teletype marked 'YES', 'NO', or '?'.

The details of this method of presentation are given in section 4.7.
During the human interviews, professional interviewers were either reading the questions (as described in section 6.5), or they were developing their own format of asking the questions of the basic questionnaire (sections 8.6 and 10.6).

3.3.6 The Semantic Differential Method

The attitude of interviewees towards computer or human interviews was assessed by using a device based on the Semantic Differential method; interviewees were asked to tick one of seven boxes (representing a rating scale) for each of six related bipolar adjectives, in order to show how much they 'liked' or they 'disliked' the interview they had.

The detailed development of the method is described in section 4.5.

The six bipolar adjectives used, which constitute a factor called Personal Evaluation, and the way they are related was decided by using the factor analysis method (described in section 4.8).

3.3.7 Analysis of Results

Several versions of the ANOVA technique were used to analyse the results of the five experiments measuring the attitude of people towards the computer or human interview. The detailed methods used are presented in sections 4.9, 6.7 and 8.8.

To analyse the results of the experiment measuring reproducibility and accuracy of information during the computer and human interviews a sophisticated statistical model was used. This model was based on the method of maximum likelihood and on a computational algorithm, and is described in detail in section 10.7.
4. AN EXPERIMENT TO DEVISE A METHOD FOR MEASURING PEOPLE'S ATTITUDE

4.1 Aims

The main purpose of this experiment was to devise, refine and finalise a method to measure people's attitude towards the interview. This device would be used in this and in future experiments, to measure the attitude of the interviewees towards several versions of the same basic questions, but with different degrees of Encouragement and Chattiness. In that way, the effect of L.S.E. on people's attitude towards computer interviews would be studied, and an optimum version could probably be decided.

A second aim of the experiment was to give some first information about the influence of L.S.E. on people's attitude towards the computer interview. This was made possible by using three versions of the same basic questionnaire, but with a different density of L.S.E. in each version.

4.2 Material

The material used during this experiment was a questionnaire about leisure time and people's hobbies. The subject was chosen to be a 'general', 'innocent' one, for which everyone would be expected to have an interest of some kind, so as not to require as subjects a specific group from among the general population.

The questionnaire consists of 25 basic questions, which require 'YES' or 'NO' answers. The number and kind of the questions, the introduction preceding them and the thanks given to the interviewee at the end of the questionnaire, are similar to the corresponding sections of the four programs studied during the previous analysis. The original questionnaire was revised with the help of members of the Institute of Consumer Ergonomics in Loughborough (Galer, M., Personal Communication), who have considerable experience with questionnaires and interviews.
Subsequently, during a pilot study using fifteen students from Loughborough University, the questionnaire was further tested and then revised to make the questions as unambiguous and easily understood as possible. Figure 29(a) and (b) shows the S.S.Q.s of the basic questionnaire together with the relevant branching instructions, while Appendix 3 shows a typical introduction and ending of a computer interview.

4.3 Variables

There were two main variables representing the L.S.E., the E.P.s (Encouragement) and the C.P.s (Chattiness).

Three levels of each variable were used. The first level contained no E.P.s or C.P.s, the second had sparse E.P.s or C.P.s, and the third level dense E.P.s or C.P.s.

The level of the sparse variable was defined as the same amount of the variable (Encouragement or Chattiness) as that of the four questionnaires, which have been analysed (i.e. 32% of the questions included E.P.s, while 24% of them included C.P.s), and the level of the dense variable as twice as many E.P.s or C.P.s as the sparse one (to compensate in a way for the 30% context-sensitive L.S.E., which were omitted).

In this way nine versions of the questionnaire could be produced, to cater for all the possible combinations of the levels of the two variables (a three by three matrix). During the described experiment only three versions of the questionnaire were used. One with no E.P.s or C.P.s, the second with sparse E.P.s and sparse C.P.s, and the third with dense E.P.s and dense C.P.s. The two variables were in this case used as a combined single variable, the corresponding experiment being a single factor experiment.
The pattern, which the two variables followed, was a random one, while their density was fixed for each version. In fact the patterns of L.S.E. were not absolutely random, but there was a slight deviation from randomness to prevent apparently unsympathetic or even stupid and distressing replies, e.g. E.P.s giving positive reinforcements as congratulating phrases should not follow a negative answer, that is an answer revealing a social disadvantage.

4.4 Subjects

The subjects used in this experiment were people living around N.P.L., London, and were intended to represent the General Public. The selection was done as follows: letters were distributed to most people in walking distance from N.P.L., asking them politely if they would be willing to participate in the experiment. From among those responding to this call, seventy-eight subjects were randomly chosen to take part in the experiment. The subjects were between fifteen and fifty-five years of age, as very young or very old subjects do not always present good reliability in their scores (Di Vesta and Dick, 1966; Maltz, 1963).

The majority of the respondents were housewives and hence the majority of the seventy-eight subjects were housewives too. However, by a balancing process it was ensured that equal numbers of men and women took part in each of the three versions of the questionnaire.

4.5 The Semantic Differential Method

As the basis for a method to measure people's attitude towards the interview the Semantic Differential method was used.

A number of twenty bipolar adjectives were selected from the existing bibliography, as the most relevant to this study. The thesaurus study (Osgood, et al. 1957), the work of R. Lucas
(Lucas, 1974), and that of N. Bevan from N.P.L. (Bevan, Personal Communication) were the sources of the bipolar adjectives among which the twenty most relevant to this work were chosen. Table 4 shows these 20 bipolar adjectives in alphabetical order:

Table 4: The Bipolar Adjectives

1) Complete-Incomplete 11) Personal-impersonal
2) Co-operative-Competitive 12) Pleasant-Unpleasant
3) Excitable-Calm 13) Positive-Negative
4) Fast-Slow 14) Relaxed-Tense
5) Friendly-Unfriendly 15) Sensitive-Insensitive
6) Graceful-Awkward 16) Sharp-Dull
7) Interesting-Boring 17) Simple-Complex
8) Kind-Cruel 18) Sophisticated-Naive
9) Natural-Artificial 19) Stable-Changeable
10) Passive-Active 20) Strong-Weak

In the experiment the single concept of 'the computer interview which the subjects had' was rated on each of the 20 bipolar adjectives. Each of the two opposite poles were presented at the end of a line marked in seven equidistant boxes, the scale being labelled over the boxes with the adverbs "slightly, quite, extremely" either side of the neutral position. The use of these adverbial quantifiers is justified on the grounds that (a) they increase the amount of differentiation and help the subjects to spread their ratings more evenly (Wells and Smith, 1960a and b); and (b) they help the interviewee to understand the labelled scales and lead to greater co-operation in the rating task (Heise, 1969). Furthermore, the results of several studies (Cliff, 1959; Howe, 1962, 1966a, 1966b) suggest that the above adverbs do define rating positions which are equidistantly spaced. This would of course facilitate the mathematical handling of the scores.
Our own experience, gained during a pilot study, was that adverbial quantifiers do help the interviewees to understand, and increase their co-operation in the rating task (comparison between two presentations one with adverbial quantifiers the other without them).

4.6 The Computer System.

The computer system used during this experiment was written in an extended version of the BASIC computer language running in a PDP 11/10 mini-computer, and worked in two stages.

During the first stage it accepted the three versions of the questionnaire and the logical structure connecting them. At this stage the system was also fed with the bipolar adjectives in alphabetical order. Each version of the questionnaire was entered as a sequence of numbered blocks. For convenience each question was represented by one block and was preceded by four numbers. The first of these numbers was the number of the question, while the other three were the logical connections of this question in the questionnaire context, that is the number of the next question for a possible 'YES' or 'NO' or '?' answer of the interviewee correspondingly. The bipolar adjectives were randomized by the system before each presentation. This randomization took place not only on the order of the bipolar adjectives, but also on their polarity (directionality).

During the second stage the system controlled the interview process (i.e. carried out the interview). During the interview the system was able, through suitable arrays, to keep trace of the presentation time of each question, and the user's reaction time for each question (i.e. the time it took the user to give a 'YES', 'NO' or '?' answer after the presentation of a question), together with the number and response of each question asked.
4.7 The Man-Computer Interaction

Each person, coming to the N.P.L. Laboratories for the computer interview, was told that the computer was going to ask him/her a few simple 'YES' or 'NO' questions about leisure time and hobbies. He/she was also told that he/she could answer the questions by pressing one of the three buttons of a keyboard labelled 'YES', 'NO', '?' (these were actually shown to the user), and that he/she had nothing to worry about the interview, because the computer would assist him/her throughout the whole process. Then the subject was left alone in a room in front of a keyboard and a V.D.U.

Before leaving the room the experimenter started the system by pressing a few buttons of the keyboard, and then the system started interacting with the user by presenting the questions in natural English on a V.D.U. at a rate of 30 characters/sec.

Each person of the 78 used had only one interview with the computer on one of the three questionnaire versions (the interviewees being randomly assigned to one of the three versions).

Then the interviewee was asked by the computer to rate the interview and the way the questions were phrased on each of the 20 bipolar adjectives using a seven point scale, the scale positions being labelled by adverbial quantifiers as already described, and was assisted in doing so by Computer Aided Instructions (C.A.I.). Accordingly before starting the rating process, the subjects were told by the computer to give their first impression and not to think of the 'correct' answer, because there was no such answer. Then the seven point scale, with the adverbial qualifiers and the two opposite poles of the bipolar adjectives were presented on the V.D.U., and examples were given by the computer of how the interviewees could 'mark' their preference on the scale. Throughout the interview process, the subjects were also asked to confirm their scores by pressing the same key of the keyboard for a second time. (The latter because the author was informed through personal communication with Mr. Bevan of N.P.L., that the subjects sometimes wish to change their first rating scores.)
It can be said that the interview itself went smoothly, but the rating process after the interview presented slight problems with about 20% of the subjects asking for human assistance.

4.8 The Factor Analysis - Results with the Semantic Differential Method

During the interview, it was obvious for some people (20% - 30% of the total) that they did not understand the rating process. This was made obvious especially when they were asked by the computer to confirm their answers, or when they were asking for human assistance, or by their scores which were almost invariably on the same point of the scale for each bipolar adjective.

Before the interview the comprehension scores of all the subjects had been assessed by the comprehension part of a Nelson-Denny Reading Test (Brown et al. 1973). Trying to remove the unreliability of the data, due to the scores of the people who did not understand the rating process, about 30% of the rating scores, those belonging to the people with lower comprehension scores, were eliminated (the total of the 20% people asking for human assistance being included in the discarded data).

The remaining rating scores of 18 people on each version were then factor analysed, to find out groups of relevant (correlated) bipolar adjectives. The Principal Components analysis was employed, and then the factors were rotated according to the Varimax method. This was done by the computer using a program called 'Little Jiffy, Mark IV' (Kaiser and Rice, 1974).
By using the criterion of $1^+$ eigenvalue, six factors emerged. These factors measured 64% of the common variance, with all significant loadings on each factor being at least 0.28. Their relative contributions were 26.10%, 23.13%, 14.92%, 12.65%, 12.03% and 11.17% correspondingly.

The six factors were given the names: 1) Personal Evaluation, 2) Potency or Functionality, 3) Social Evaluation, 4) Simplicity, 5) Activity and 6) Stability. Table 5 presents the six factors and the bipolar adjectives contributing to each factor. The number in parentheses on the side of each factor is an index of the factor's reproducibility, while the numbers in parentheses by the side of each bipolar adjective are the factor loadings.

The first factor (Personal Evaluation) was chosen to measure the attitude of the people towards the interview (in this and the subsequent four experiments), because it contains all the pairs of adjectives which are relevant to the variables to be measured. It is noticeable that the loadings of the six bipolar adjectives which constitute the first factor (representing the common variance between each adjective and the factor), are at least 0.51 for each bipolar adjective.

Therefore it can be said as a conclusion of this section that the Semantic Differential method and the Factor Analysis statistical technique were combined in this experiment to produce the final method to measure people's attitude towards the interview. This method consists of six linearly related bipolar adjectives, which constitute the factor Personal Evaluation, and was used for all the experiments of this research, to measure people's attitude towards the interview.
Table 5: The Factors

1. **PERSONAL EVALUATION (0.73)**
   - Kind-Cruel (0.79)
   - Graceful-Awkward (0.73)
   - Pleasant-Unpleasant (0.69)
   - Interesting-Boring (0.60)
   - Relaxed-Tense (0.53)
   - Friendly-Unfriendly (0.51)

2. **POTENCY OR FUNCTIONALITY (0.68)**
   - Positive-Negative (0.86)
   - Complete-Incomplete (0.60)
   - Personal-Impersonal (0.59)
   - Sharp-Dull (0.56)
   - Passive-Active (-0.46)
   - Interesting-Boring (0.45)
   - Strong-Weak (0.42)
   - Natural-Artificial (0.42)
   - Friendly-Unfriendly (0.39)

3. **SOCIAL EVALUATION (0.65)**
   - Co-operative-Competitive (0.75)
   - Excitable-Calm (-0.65)
   - Sensitive-Insensitive (0.59)
   - Personal-Impersonal (0.40)
   - Natural-Artificial (0.39)

4. **SIMPLICITY (0.60)**
   - Simple-Complex (0.81)
   - Relaxed-Tense (0.56)
   - Natural-Artificial (0.40)
   - Sensitive-Insensitive (-0.36)
   - Excitable-Calm (0.30)

5. **ACTIVITY (0.49)**
   - Fast-Slow (0.89)
   - Passive-Active (-0.58)
   - Sophisticated-Naive (0.40)
   - Sharp-Dull (0.29)

6. **STABILITY (0.50)**
   - Stable-Changable (0.86)
   - Sophisticated-Naive (0.43)
   - Personal-Impersonal (0.36)
   - Excitable-Calm (-0.30)
   - Complete-Incomplete (-0.28)
4.9 Further Analysis of the Results - The Questionnaire Versions

The fact that three versions of the same basic questionnaire, but with different amount of E.P.s and C.P.s, were used made it possible to have some first information about the effect of the L.S.E. on the attitude of the people towards the computer interview.

Accordingly the ratings of each version were compared for significant differences, and this was done for the six factors and for all the six bipolar adjectives of the first factor.

It is believed that the rating scores of the people on each bipolar adjective express and measure their attitude towards three things:

a) the meaning of the questions (interest in the subject matter);
b) the environment during the interview (i.e. mainly the media used - computer, screen, keyboard, method of presenting the questions etc.), and

c) the style of the questions or the way the questions were phrased.

This and the fact that each person used only one version of the questionnaire among the three, for obvious reasons (between- rather than within- subjects design), made it more difficult to have significant results, though significant results, if any were found, would be reliable.

In any case the analysis of variance for a single factor experiment was employed and a linear model was assumed.

The structural model can be represented mathematically by the following equation: $X_{ij} = U + t_j + e_{ij}$ where

$X_{ij}$ = the criterion measure on a randomly selected element $i$ in the treatment population $j$

$U$ = grand mean of treatment populations

t$ = effect for treatment $j$ ($t_j = U_j - U$ where $U_j$ = the mean of treatment $j$)

e$ = experimental error.
The model's factor (experimental variable) was considered a fixed one for this experiment, because its levels were determined by a systematic, non-random procedure (Winer, 1962, page 313). For this case \( u \) is constant for all measurements in all treatment populations, \( t_j \) is constant for all measurements within population \( j \) but not for population \( k \), where \( k \) represents some treatment other than \( j \). The experimental error \( e_{ij} \) represents all the uncontrolled sources of variance affecting individual measurements, with unique effect for each element \( i \) in the basic population. This effect is further assumed to be independent of \( t_j \).

For this experiment \( j = 1, 2, 3 \) represents the treatments associated with the three levels of the factor, \( i = 1, 2, \ldots, 18 \) represents the number of people who took part in every treatment, and \( X_{ij} \) the rating score of person \( i \) for version \( j \) of the questionnaire on one of the bipolar adjectives.

Tables 6, 7 and 8 show the scores of the 54 interviewees on each of the six bipolar adjectives of Factor 1, while the averages and standard variations of these scores are given in Table 9 (standardised and multiplied by 100). Figure 30 shows the averages and standard deviations of the three versions of the questionnaire for the first factor. The factor scores have been calculated from the raw scores according to the method of composite estimates (Rummel, 1970, page 441).

The ANOVA of those results presented an overall significance at the 0.05 level for the first factor, and each of the six bipolar adjectives of the first factor. Further analysis showed that this was mainly attributable to the difference between the second version and the other two versions combined, while versions 1 and 3 showed no significant differences.
Table 6: Scores for the first two Adjectives of Factor 1  
(Attitude scores - each Version/Row No. by a different subject.)

a) Kind-Cruel

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

b) Graceful-Awkward

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>16</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Row No.</td>
<td>Vers.1</td>
<td>Vers.2</td>
<td>Vers.3</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>-2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 8: The Fifth and Sixth Adjectives of Factor 1
(Attitude rating scores - each Version/Row No. by a different subject.)

e) Relaxed-Tense  f) Friendly-Unfriendly

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>-1</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td>-1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>13</td>
<td>-2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>-1</td>
<td>3</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>-1</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 9: The scores on each Bipolar Adjective of Factor 1

<table>
<thead>
<tr>
<th>Bipolar Adjective</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Standard Deviation Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kind-Cruel</td>
<td>0.60</td>
<td>1.60</td>
<td>1.10</td>
<td>0.91</td>
<td>1.93</td>
<td>1.23</td>
</tr>
<tr>
<td>Graceful-Awkward</td>
<td>0.10</td>
<td>0.80</td>
<td>-0.20</td>
<td>1.13</td>
<td>1.15</td>
<td>1.31</td>
</tr>
<tr>
<td>Pleasant-Unpleasant</td>
<td>1.11</td>
<td>2.05</td>
<td>1.27</td>
<td>1.32</td>
<td>0.87</td>
<td>1.07</td>
</tr>
<tr>
<td>Interesting-Boring</td>
<td>0.30</td>
<td>1.80</td>
<td>0.60</td>
<td>1.78</td>
<td>0.92</td>
<td>1.57</td>
</tr>
<tr>
<td>Relaxed-Tense</td>
<td>1.50</td>
<td>2.61</td>
<td>1.39</td>
<td>1.46</td>
<td>0.50</td>
<td>1.65</td>
</tr>
<tr>
<td>Friendly-Unfriendly</td>
<td>0.61</td>
<td>1.83</td>
<td>0.88</td>
<td>1.45</td>
<td>0.85</td>
<td>1.27</td>
</tr>
</tbody>
</table>

The remaining 5 factors presented no significant difference, apart from factor 5 the Activity factor. This factor presented overall significant results at the level of 0.05, with version 2 perceived by the people as significantly more active than version 3, and version 3 than version 1.

Tables 10 and 11 show the total reaction time of each interviewee, and the total interaction time correspondingly. These scores show that the average reaction time for version 2 was longer, with version 3 following, but this difference was not significant. (This result is in complete agreement with the findings on the reaction time per question.) Significant results were found, at the level of 0.05, for the total interaction time spent for the interviews of each version, with version 3 having a higher average than version 2 and version 2 than version 1. The significant results on the total interaction time were also confirmed by similar findings of the interaction time per question, which was significantly different for the three versions too in a similar way to that of the total interaction time.

All the above results could be considered to present some uncertainty, because the normal distribution of the data, which is an important requirement of the analysis of variance, was rather difficult to be satisfied for this kind of data (discrete rather than continuous); therefore the results were also checked by a non-parametric method. Accordingly the Kruskal-Wallis test was employed but showed no difference from the previous results (i.e. the differences reported as significant were also significant by the Kruskal-Wallis test at the same significance levels.)
### Table 10: Reaction Time (in seconds)

(Each Version/Row No. refers to a different subject.)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92.3</td>
<td>248.2</td>
<td>46.9</td>
</tr>
<tr>
<td>2</td>
<td>122.8</td>
<td>171.5</td>
<td>84.1</td>
</tr>
<tr>
<td>3</td>
<td>109.2</td>
<td>148.6</td>
<td>204</td>
</tr>
<tr>
<td>4</td>
<td>47.7</td>
<td>69.5</td>
<td>114.1</td>
</tr>
<tr>
<td>5</td>
<td>154.8</td>
<td>53.8</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>60.1</td>
<td>128.8</td>
<td>91.1</td>
</tr>
<tr>
<td>7</td>
<td>103.2</td>
<td>183.9</td>
<td>189.1</td>
</tr>
<tr>
<td>8</td>
<td>48</td>
<td>90</td>
<td>59.8</td>
</tr>
<tr>
<td>9</td>
<td>68.6</td>
<td>102.4</td>
<td>63.7</td>
</tr>
<tr>
<td>10</td>
<td>103</td>
<td>249.4</td>
<td>140.1</td>
</tr>
<tr>
<td>11</td>
<td>83.6</td>
<td>203.5</td>
<td>151.1</td>
</tr>
<tr>
<td>12</td>
<td>256.4</td>
<td>68.2</td>
<td>153.4</td>
</tr>
<tr>
<td>13</td>
<td>88.3</td>
<td>178.5</td>
<td>99.8</td>
</tr>
<tr>
<td>14</td>
<td>86.3</td>
<td>259.3</td>
<td>231.2</td>
</tr>
<tr>
<td>15</td>
<td>114.8</td>
<td>103.9</td>
<td>155.8</td>
</tr>
<tr>
<td>16</td>
<td>167.7</td>
<td>61.6</td>
<td>121.3</td>
</tr>
<tr>
<td>17</td>
<td>142.7</td>
<td>229.7</td>
<td>166.8</td>
</tr>
<tr>
<td>18</td>
<td>99.1</td>
<td>72.3</td>
<td>110.9</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\Sigma &= 1948.6 \quad 2623.1 \quad 2272.2 \\
\bar{X} &= 108.3 \quad 145.7 \quad 126.2 \\
\text{S.D.} &= 49.8 \quad 71.7 \quad 51.6
\end{align*}
\]

(per subject) \(\bar{X} = 108.3\) \(145.7\) \(126.2\) \(\text{S.D.} = 49.8\) \(71.7\) \(51.6\)

### Table 11: Total Interaction Time (in seconds)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>274.4</td>
<td>452</td>
<td>263.1</td>
</tr>
<tr>
<td>2</td>
<td>305.1</td>
<td>369.3</td>
<td>261.5</td>
</tr>
<tr>
<td>3</td>
<td>303</td>
<td>340.3</td>
<td>416.3</td>
</tr>
<tr>
<td>4</td>
<td>221</td>
<td>265</td>
<td>312.3</td>
</tr>
<tr>
<td>5</td>
<td>339.9</td>
<td>245.2</td>
<td>305.3</td>
</tr>
<tr>
<td>6</td>
<td>238.1</td>
<td>319.7</td>
<td>309.1</td>
</tr>
<tr>
<td>7</td>
<td>279.2</td>
<td>365.6</td>
<td>396.9</td>
</tr>
<tr>
<td>8</td>
<td>224</td>
<td>285.7</td>
<td>280.5</td>
</tr>
<tr>
<td>9</td>
<td>244.6</td>
<td>302.3</td>
<td>273.8</td>
</tr>
<tr>
<td>10</td>
<td>279</td>
<td>442.1</td>
<td>354.6</td>
</tr>
<tr>
<td>11</td>
<td>269.4</td>
<td>396.6</td>
<td>373.5</td>
</tr>
<tr>
<td>12</td>
<td>434.6</td>
<td>256</td>
<td>353</td>
</tr>
<tr>
<td>13</td>
<td>272.3</td>
<td>347.7</td>
<td>321.3</td>
</tr>
<tr>
<td>14</td>
<td>270.2</td>
<td>450.3</td>
<td>448</td>
</tr>
<tr>
<td>15</td>
<td>297.5</td>
<td>308.2</td>
<td>370</td>
</tr>
<tr>
<td>16</td>
<td>349.7</td>
<td>267.6</td>
<td>350.1</td>
</tr>
<tr>
<td>17</td>
<td>327.9</td>
<td>422.2</td>
<td>377.3</td>
</tr>
<tr>
<td>18</td>
<td>282.9</td>
<td>262</td>
<td>330.1</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\Sigma &= 5212.8 \quad 6097.8 \quad 6196.7 \\
\bar{X} &= 289.6 \quad 338.8 \quad 344.3 \\
\text{S.D.} &= 51.1 \quad 70.9 \quad 34.6
\end{align*}
\]

(per subject) \(\bar{X} = 289.6\) \(338.8\) \(344.3\) \(\text{S.D.} = 51.1\) \(70.9\) \(34.6\)

Total numbers of questions = 28 (including introductory questions)

\[
\begin{align*}
\bar{X} &= 4.4 \quad 6.0 \quad 4.9 \\
\text{S.D.} &= 2.1 \quad 3.5 \quad 2.4
\end{align*}
\]

(per subject) \(\bar{X} = 4.4\) \(6.0\) \(4.9\) \(\text{S.D.} = 2.1\) \(3.5\) \(2.4\)

(per question) \(\bar{X} = 11.7\) \(13.9\) \(14.7\) \(\text{S.D.} = 4.9\) \(5.4\) \(5.6\)
4.10 Conclusions

The first experiment has finalised the method to measure people's attitude towards the interview in this and all subsequent relevant experiments. Moreover it has indicated that the context free L.S.E. (E.P.s and C.P.s) have a significant effect on people's acceptance of computer interviews. Finally this experiment has shown that there appears to be an optimum in the amount of random E.P.s and random C.P.s used in a computer interview.
5. AN EXPERIMENT TO COMPARE SEVERAL RANDOM PATTERNS OF E.P.s

5.1 Aims

The positive results of the first experiment (i.e. significant effect of E.P.s and C.P.s on interviewee's acceptance of computer interviews, and indication of an optimum amount of E.P.s and C.P.s) encouraged a second experiment. The aim of this experiment was to examine in more detail the effect of the variable Encouragement only, on the acceptability of computer interviews to people. (In experiment 1 the combined effect of both variables was examined.)

5.2 Material

The same questionnaire about leisure time and people's hobbies, which was used for the first experiment, was employed for this experiment too. (The basic questionnaire and the relevant branching instructions are shown in figure 29 (a) and (b).)

5.3 Variables

The variable Encouragement (E.P.s) was the only controlled variable of this experiment.

Three levels of this variable were used. The first with no amount of the variable, the second with sparse E.P.s, and the third with dense E.P.s. Here again, as with the first experiment, the level of the sparse variable was defined as the same amount of this variable as that of the four questionnaires, which have been analysed in Chapter 2, and the level of the dense variable as twice as many E.P.s as the sparse one.

Consequently during this experiment three versions of the questionnaire, about leisure time and people's hobbies, were used corresponding to the three levels of the variable Encouragement, which were mentioned above.

Again the pattern of the E.P.s was a random one with fixed density for each version, as the previous analysis has indicated. The same slight deviation from randomness of the first experiment applied to this experiment too (i.e. E.P.s giving positive
reinforcements, as congratulating phrases, should not follow
an answer of the interviewee showing a personal failure or a
social disadvantage and vice-versa).

5.4 Subjects
The subjects used in this experiment were students and
young apprentices of the University and the Colleges of
Loughborough.

The selection of the subjects was done as follows: students
were approached at random in the campus of the University, and were
asked politely to participate in the experiment. The first 60,
of those who gave a positive response (refusals about 30%), were
chosen to take part in the experiment i.e. twenty per each of the
three versions of the questionnaire.

More men than women took part in the experiment, but due
to a balancing process, equal number of women answered each of
the three versions.

5.5 The Computer System
The computer system for that experiment was written in the
BASIC-PLUS computer language, slightly different from that of
experiment 1, running in a PDP 11/40 mini-computer instead of
a PDP 11/10 one, and worked in two stages.

During the first stage the system accepted the three versions
of the questionnaire, and the logical structure connecting them
(i.e. the number of the next question after a possible 'YES', 'NO'
or '?' answer of the interviewee). The questionnaire was entered
as a sequence of numbered blocks, each question represented by a
block, while three arrays were used in order to store the logical
structure connecting the questions.
During the second stage the system carried out the interview. During the interview the system was able, by using suitable vector arrays, to keep trace of the number and the response of each question asked, the user's reaction time for each question (i.e. the time it took the user to give a 'YES', 'NO' or '?' answer), and the time of presentation of each question.

The system used for that experiment was written by the author and a list of it is given in Appendix 2.

Rules for the construction of the software system for any environment (independent of the computer language used) are given in Chapter 12.)

5.6 The Man-Computer Interaction

Each person, coming to the Computer Studies Laboratories for the experiment, was told that the computer was going to ask him/her a few simple 'YES' or 'NO' questions about leisure time and hobbies. He/she was also told that he/she could answer the questions by pressing one of the three buttons of a keyboard labelled 'YES', 'NO', '?' (these were actually shown to the user), the '?' to be used if a question was not comprehended. Then each subject was told that he/she had nothing to worry about the interview, because the computer would assist him/her throughout the whole process, and was left alone in a room in front of a keyboard and a V.D.U.

Before leaving the room the experimenter initialised the system and then the system started interacting with the user, by presenting the questions in natural English on the V.D.U. at a rate $> 100$ characters/sec.

Each of the sixty subjects had only one interview with the computer on one of the three questionnaire versions, the interviewees being randomly assigned to one of the three versions. This interaction went on smoothly and presented no problem for all sixty subjects.
5.7 The Semantic Differential Method

As for the first experiment, the S.D. method was used to measure people's attitude towards the interview.

The six bipolar adjectives of Factor 1, devised from experiment 1, were used.

In this experiment the subjects were asked to rate the single concept of 'the computer interview which they had', on each of the six bipolar adjectives using paper and pencil. Each of the two opposite poles of the adjectives was presented at the end of a part of a line divided into seven equidistant boxes, the scale positions labelled with the adverbs slightly, quite, extremely, around the neutral position, as shown in figure 31. (The adjectives 'GOOD-BAD' were used by the experimenter as an example, in order to show the subjects how to use the S.D. scale.)

Moreover four different forms of presenting the six adjectives were used throughout the experiment, differing only in the order and directionality of the adjectives. The order and polarity of the adjectives of each version was random, and so was the process by which one of the four versions was allocated to each subject.

The question of a difference in the reliability of the attitude scores of the first experiment and this experiment should also be considered at this stage. This is because in the first experiment the six bipolar adjectives of factor 1 were presented with 14 more adjectives, while in this experiment we had 6 bipolar adjectives only. Although an interaction of the 6 bipolar adjectives with the remaining 14 in experiment 1 does not seem very likely, it would be reasonable to compare the scores on the six bipolar adjectives of this experiment with those of experiment 1 to find out if the data show such an interaction or not (possibly for Version 1 of the interview questionnaire, which is common in both experiments.)
5.8 Analysis of Results

The ratings of the three versions of the questionnaire, each with different degree of the variable encouragement, were compared for significant differences for all six adjectives of Factor 1.

Tables 12, 13 and 14 show the scores of 20 students per version of the questionnaire, while Table 15 gives their averages and standard deviations for each of the six bipolar adjectives. As for experiment 1 each person used only one version of the questionnaire among the three (between- rather than within-subjects design).

Again the analysis of variance for a single factor experiment was employed, while the model's experimental variable was considered a fixed one as for experiment 1, for similar reasons. The structural model, which the analysis of variance was based on, was exactly the same to that presented in experiment 1.

The combined factors scores for each of the three versions were again calculated according to the method of composite estimates and are shown in figure 32. These scores presented results significant at least at the 0.05 level for each of the six bipolar adjectives, while for three out of the six adjectives and for the combined factor scores the results were significant at the 0.01 level.

Further analysis showed that this difference was mainly attributable to the difference of version 2 and the other two versions combined.

Moreover by comparing the combined factor scores for version 1 of this experiment with the combined factor scores for the same version of experiment 1 no significant differences were found (analysis of variance, p > 0.05). The little difference of the combined factor scores of version 1 for the two experiments can also be observed in figures 30 and 32. The above results indicate that there was no significant interaction effect between the 6 bipolar adjectives of factor 1 and the remaining 14 bipolar adjectives of experiment 1. This therefore suggests that the factor 1 set of 6 bipolar adjectives may be accepted as a reliable basis for assessing attitudes to these interviews.
Table 12: Experiment 2: First and Second Adjectives of Factor 1

(Attitude rating scores - each Version/Row No. by a different subject.)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>-2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>-2</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>-1</td>
<td>-3</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>-1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>18</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>19</td>
<td>-1</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
</tr>
</tbody>
</table>
Table 13: Experiment 2: Third and Fourth Adjectives of Factor 1
(Attitude scores - each Version/Row No. by a different subject.)
c) Pleasant-Unpleasant
d) Interesting-Boring

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>2</td>
<td>-1</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>16</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 14: Experiment 2: Fifth and Sixth Adjectives of Factor 1
(Attitude scores - each Version/Row No. by a different subject.)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>-3</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>15</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>19</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>3</td>
<td>-2</td>
<td>20</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 15: Experiment 2: The Scores on each Bipolar Adjective of Factor 1

<table>
<thead>
<tr>
<th>Bipolar Adjectives</th>
<th>Average Vers.1</th>
<th>Average Vers.2</th>
<th>Average Vers.3</th>
<th>Standard Deviation Vers.1</th>
<th>Standard Deviation Vers.2</th>
<th>Standard Deviation Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kind-Cruel</td>
<td>1.00</td>
<td>2.05</td>
<td>1.15</td>
<td>1.03</td>
<td>1.05</td>
<td>1.35</td>
</tr>
<tr>
<td>Graceful-Awkward</td>
<td>-0.30</td>
<td>1.35</td>
<td>-0.35</td>
<td>1.22</td>
<td>1.50</td>
<td>1.27</td>
</tr>
<tr>
<td>Pleasant-Unpleasant</td>
<td>1.70</td>
<td>2.30</td>
<td>1.40</td>
<td>0.73</td>
<td>0.47</td>
<td>1.44</td>
</tr>
<tr>
<td>Interesting-Boring</td>
<td>1.55</td>
<td>2.30</td>
<td>0.85</td>
<td>0.89</td>
<td>0.66</td>
<td>1.53</td>
</tr>
<tr>
<td>Relaxed-Tense</td>
<td>1.85</td>
<td>2.50</td>
<td>1.65</td>
<td>1.40</td>
<td>0.51</td>
<td>2.09</td>
</tr>
<tr>
<td>Friendly-Unfriendly</td>
<td>2.00</td>
<td>2.55</td>
<td>1.55</td>
<td>0.92</td>
<td>0.51</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Tables 16 and 17 show the total reaction time and the total overall interaction time for each interviewee correspondingly. The average reaction time of version 2 was again longer, but these results did not present an overall significant difference. The results of the total interaction time did not present an overall significant difference too. (Similar results were also found by analysing the reaction time per question and the total interaction time per question.)

5.9 Conclusions

This experiment has shown that the E.P.s above, when following a random pattern among the questions, may have a significant effect on people's acceptance of computer interviews. It has also shown that there appears to be an optimum in the amount of E.P.s used in a computer interview.

The above results are in complete agreement with those of experiment 1, and both experiments confirm the usefulness of the context-free L.S.E. in a computer interview.
<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>123</td>
<td>177</td>
<td>148</td>
<td>1</td>
<td>157</td>
<td>207</td>
<td>138</td>
</tr>
<tr>
<td>2</td>
<td>209</td>
<td>142</td>
<td>103</td>
<td>2</td>
<td>249</td>
<td>180</td>
<td>197</td>
</tr>
<tr>
<td>3</td>
<td>185</td>
<td>146</td>
<td>279</td>
<td>3</td>
<td>220</td>
<td>186</td>
<td>195</td>
</tr>
<tr>
<td>4</td>
<td>173</td>
<td>166</td>
<td>150</td>
<td>4</td>
<td>202</td>
<td>229</td>
<td>138</td>
</tr>
<tr>
<td>5</td>
<td>87</td>
<td>168</td>
<td>196</td>
<td>5</td>
<td>118</td>
<td>233</td>
<td>320</td>
</tr>
<tr>
<td>6</td>
<td>121</td>
<td>115</td>
<td>115</td>
<td>6</td>
<td>163</td>
<td>146</td>
<td>192</td>
</tr>
<tr>
<td>7</td>
<td>108</td>
<td>153</td>
<td>130</td>
<td>7</td>
<td>142</td>
<td>186</td>
<td>244</td>
</tr>
<tr>
<td>8</td>
<td>110</td>
<td>244</td>
<td>150</td>
<td>8</td>
<td>172</td>
<td>185</td>
<td>163</td>
</tr>
<tr>
<td>9</td>
<td>123</td>
<td>169</td>
<td>148</td>
<td>9</td>
<td>152</td>
<td>214</td>
<td>163</td>
</tr>
<tr>
<td>10</td>
<td>125</td>
<td>146</td>
<td>149</td>
<td>10</td>
<td>240</td>
<td>217</td>
<td>181</td>
</tr>
<tr>
<td>11</td>
<td>210</td>
<td>155</td>
<td>182</td>
<td>11</td>
<td>263</td>
<td>184</td>
<td>187</td>
</tr>
<tr>
<td>12</td>
<td>221</td>
<td>118</td>
<td>157</td>
<td>12</td>
<td>272</td>
<td>191</td>
<td>206</td>
</tr>
<tr>
<td>13</td>
<td>220</td>
<td>162</td>
<td>186</td>
<td>13</td>
<td>224</td>
<td>266</td>
<td>283</td>
</tr>
<tr>
<td>14</td>
<td>164</td>
<td>148</td>
<td>107</td>
<td>14</td>
<td>187</td>
<td>205</td>
<td>232</td>
</tr>
<tr>
<td>15</td>
<td>151</td>
<td>184</td>
<td>140</td>
<td>15</td>
<td>232</td>
<td>178</td>
<td>196</td>
</tr>
<tr>
<td>16</td>
<td>107</td>
<td>215</td>
<td>148</td>
<td>16</td>
<td>239</td>
<td>220</td>
<td>239</td>
</tr>
<tr>
<td>17</td>
<td>205</td>
<td>236</td>
<td>143</td>
<td>17</td>
<td>218</td>
<td>257</td>
<td>147</td>
</tr>
<tr>
<td>18</td>
<td>174</td>
<td>183</td>
<td>163</td>
<td>18</td>
<td>192</td>
<td>162</td>
<td>169</td>
</tr>
<tr>
<td>19</td>
<td>151</td>
<td>173</td>
<td>104</td>
<td>19</td>
<td>188</td>
<td>214</td>
<td>196</td>
</tr>
<tr>
<td>20</td>
<td>151</td>
<td>166</td>
<td>162</td>
<td>20</td>
<td>242</td>
<td>210</td>
<td>179</td>
</tr>
</tbody>
</table>

\[
\sum = 3118 \quad 3366 \quad 3060 \quad \sum = 3875 \quad 3970 \quad 3865
\]
(per subject) \(\bar{x} = 155.9 \quad 168.3 \quad 153\)
S.D. = 42.3 \quad 33.2 \quad 39.2

\[
\bar{x} = 193.8 \quad 198.5 \quad 193.2
\]
S.D. = 45.6 \quad 36.9 \quad 42.1

Total number of questions = 28 (including introductory questions)
(per subject \(\bar{x} = 6.1 \quad 7.5 \quad 7.0\)
and per question) S.D. = 3.1 \quad 2.6 \quad 2.6
\(\bar{x} = 8.2 \quad 8.9 \quad 9.0\)
S.D. = 2.9 \quad 3.6 \quad 3.1
6. AN EXPERIMENT TO COMPARE SEVERAL RANDOM PATTERNS OF L.S.E.
IN A HUMAN INTERVIEW

6.1 Aims

Experiments 1 and 2 have indicated that: a) random Encouragement and Chattiness have a significant effect on the acceptability of computer interviews to the interviewee, and b) there appears to be an optimum for the amount of Encouragement and Chattiness used in a computer interview.

Yet, it would be very interesting to examine the effect of the same random patterns of E.P.s and C.P.s on user's acceptance of the interview in a man-man interviewing situation, and to compare the results with those of man-computer interviewing situation. Therefore a third experiment was conducted with the above aim.

6.2 Material

The same basic questionnaire about leisure time and people's hobbies, used in experiments 1 and 2, was used in this experiment too.

6.3 Variables

The same two variables, Encouragement (E.P.s) and Chattiness (C.P.s) were used for this experiment.

Nine versions of the same basic questionnaire were generated, by using all the possible combinations of the three levels of each of the two variables. (The levels of the variables as in experiments 1 and 2.)

Yet, a small number of C.P.s were slightly changed to be adapted to the human interview e.g. the post-phrase 'Please push 'YES' or 'NO'' was changed to 'Please answer 'YES' or 'NO' etc.
The pattern of the variables was again random with fixed density for each version. However, as with experiments 1 and 2, care was taken to avoid positive reinforcements, e.g. a congratulating phrase, after answers of the interviewee which revealed a social disadvantage and vice-versa.

6.4 Subjects

The subjects used in this experiment were people of Loughborough and Leicester, and they (are supposed to) represent the General Public.

The selection of people was done as follows: the professional interviewers, who carried out the interviews, standing at suitable busy places chose people at random from among those passing by and asked them politely to participate in the experiment. The first 180 of those who gave a positive response to the call of the interviewers took part in the experiment, 20 people for each version of the questionnaire.

The subjects were between twenty and fifty years of age for the same reasons given in experiment 1.

Moreover, due to a balancing process, there was an equal number of men and women who took part in each of the nine versions of the questionnaire.

The five professional interviewers, who carried out the interview, were all women and did equal number of men and women for each of the nine versions of the interview questionnaire, i.e. 2 men and 2 women per interviewer and per questionnaire version.

6.5 The Man-Man Interaction

The subjects were told by the professional interviewers that they were going to answer a few simple YES-NO questions about leisure time and people's hobbies. They were also told that this was part of a survey on this subject, conducted by the Department of Human Sciences of Loughborough University.
The directions to the professional interviewers were to use the exact wording of the questions of each version of the questionnaire, the point being properly emphasized.

All the 5 interviewers reported that the interviews went on smoothly, and that the questions of the questionnaire seemed completely understood by the people. They also reported refusals to participate at the rate of 20-30% of the people approached. On the style of the questions, the interviewers reported that they used the exact wording of the questions and phrases, assigning randomly each subject to one of the nine versions of the questionnaire as originally instructed.

Since there was no control over the action of the 5 professional interviewers, an effort was made to check their data for sharp deviations from expected values.

Consequently the same basic questionnaire about leisure time and hobbies was sent by post to people in the Loughborough area, chosen randomly from the electoral register (refusals about 40%). The frequencies of 'YES', 'NO' and '?' responses of all 25 questions of the five interviewers and the paper and pencil questionnaire are given in tables 18, 19 and 20 correspondingly.

Based on tables 18, 19 and 20 the number of significant disagreements between each professional interviewer and the paper and pencil questionnaire was calculated, for the 25 questions of the questionnaire as shown in table 21 ($X^2$ test for each question, disagreements significant at the 0.05 level).
<table>
<thead>
<tr>
<th>Question No.</th>
<th>Int.1</th>
<th>Int.2</th>
<th>Int.3</th>
<th>Int.4</th>
<th>Int.5</th>
<th>Pencil and Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>30</td>
<td>29</td>
<td>32</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>30</td>
<td>26</td>
<td>31</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>19</td>
<td>20</td>
<td>9</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>13</td>
<td>15</td>
<td>7</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>7</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>19</td>
<td>24</td>
<td>16</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>12</td>
<td>15</td>
<td>7</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>27</td>
<td>32</td>
<td>26</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>34</td>
<td>26</td>
<td>33</td>
<td>33</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>19</td>
<td>29</td>
<td>27</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>13</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>16</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>16</td>
<td>20</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>22</td>
<td>9</td>
<td>20</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>23</td>
<td>18</td>
<td>23</td>
<td>17</td>
<td>22</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>15</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Question No.</td>
<td>Int.1</td>
<td>Int.2</td>
<td>Int.3</td>
<td>Int.4</td>
<td>Int.5</td>
<td>Pencil and Paper</td>
</tr>
<tr>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------------------</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>17</td>
<td>16</td>
<td>26</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>20</td>
<td>13</td>
<td>28</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>17</td>
<td>12</td>
<td>20</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>19</td>
<td>16</td>
<td>23</td>
<td>18</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>26</td>
<td>20</td>
<td>30</td>
<td>27</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>13</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>15</td>
<td>13</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>18</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>20</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>21</td>
<td>26</td>
<td>20</td>
<td>16</td>
<td>26</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>18</td>
<td>13</td>
<td>19</td>
<td>14</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>14</td>
<td>7</td>
<td>15</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>
## Table 20: Frequencies of '?' Answers

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Int.1</th>
<th>Int.2</th>
<th>Int.3</th>
<th>Int.4</th>
<th>Int.5</th>
<th>Pencil and Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>19</td>
<td>16</td>
<td>27</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>17</td>
<td>16</td>
<td>27</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>13</td>
<td>15</td>
<td>6</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>17</td>
<td>12</td>
<td>19</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>17</td>
<td>12</td>
<td>20</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>22</td>
<td>20</td>
<td>16</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>17</td>
<td>22</td>
<td>19</td>
<td>18</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>26</td>
<td>26</td>
<td>27</td>
<td>20</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>14</td>
<td>8</td>
<td>21</td>
<td>10</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>18</td>
<td>12</td>
<td>19</td>
<td>14</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>18</td>
<td>13</td>
<td>19</td>
<td>14</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>
Based on the results of Table 21, the following two hypotheses were tested:

1) The number of agreements and disagreements between each professional interviewer and the postal questionnaire is a random event.

2) The 5 professional interviewers agree (or disagree) with the postal questionnaire in a uniform way.

The first hypothesis (null hypothesis) was rejected for all five interviewers at the level of 0.05 (X² test), in favour of the hypothesis that an agreement is more likely than a disagreement. In fact, the null hypothesis was rejected at the 0.01 level for 4 out of the 5 professional interviewers.

The second hypothesis could not be rejected at the 0.05 level of significance (X² test).

Therefore, by considering the answers of the paper and pencil questionnaire as a rough index of the distribution of responses of the people of Loughborough area to the questions of the questionnaire, the above described analysis has shown that the interviewers' data do not deviate sharply from the expected answers.
6.6 The Semantic Differential Method

As for the previous two experiments, the attitude of people towards the interview was assessed by using the semantic differential method.

Four versions of the six bipolar adjectives of factor 1, differing in the order and polarity of the adjectives, were used, exactly as in experiment 2.

Moreover the professional interviewers were instructed to try to disassociate themselves from the interview, when asking the people to tick one of the seven boxes for each of the six bipolar adjectives after the interview. Accordingly the interviewers asked the people to rate how they liked the wording and style of the questions and not the interviewers themselves.

6.7 Analysis of Results

For statistical analysis purposes, the third experiment was a 3 x 3 factorial experiment with fixed factors.

The two factors (experimental or controlled variables), which for this particular experiment were the variables Encouragement and Chattiness, were considered fixed, because the selection of their levels (3 for each factor) was determined by a systematic, non-random procedure. Moreover, there were 20 observations per each of the nine cells, each cell corresponding to each of the nine versions of the questionnaire.

<table>
<thead>
<tr>
<th>Encouragement Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chattiness Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Note that the three versions of experiment 2 correspond to cells 1, 2 and 3 here, and the three versions of experiment 1 correspond to cells (i.e. versions) 1, 5 and 9 here.
Tables 22-27 show the scores of 180 people on the six bipolar adjectives for all nine versions of the questionnaire. The analysis of variance of these data showed no significant effect of the variable Chattiness for all six adjectives, and for the combined factor scores ($p > 0.05$). However the variable Encouragement, although it did not present significant effects for the combined factor scores, showed significant results for three out of the six bipolar adjectives at the 0.05 level. These three bipolar adjectives, namely Kind-Cruel, Graceful-Awkward, and Friendly-Unfriendly, indicated that the variable Encouragement, if following a random pattern, might affect the human interview in such a way that more Encouragement would result in less accepted versions. In general the ANOVA indicated no interaction effects between the two experimental variables.

Figure 33 shows the averages and standard deviations of the combined factor scores of the nine versions of the questionnaire used in this experiment.

6.8 Conclusions

This experiment has shown that the context-free L.S.E., when following a random pattern among the questions, have no effect on people's acceptance of human interviews. Moreover it has shown that random Encouragement might even have a negative effect, with more random Encouragement resulting in less acceptable interaction. (This result was further reinforced by the spontaneous comments of the professional interviewers, according to which the versions of high E.P.s and C.P.s seemed 'a bit unnatural or unhuman'.)

The above results are in obvious disagreement with those of the previous two experiments, and show how differently random Encouragement and Chattiness affect people's attitude towards computer interviews on the one hand, and human interviews on the other hand.
Table 22: Experiment 3 (Attitude scores - each Version/Row No. by a different subject)

a) Kind-Cruel

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Vers.4</th>
<th>Vers.5</th>
<th>Vers.6</th>
<th>Vers.7</th>
<th>Vers.8</th>
<th>Vers.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 23: Experiment 3

b) Graceful-Awkward

(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Vers.4</th>
<th>Vers.5</th>
<th>Vers.6</th>
<th>Vers.7</th>
<th>Vers.8</th>
<th>Vers.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>-2</td>
<td>-1</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
<td>3</td>
<td>-1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>-3</td>
<td>-2</td>
<td>2</td>
<td>-1</td>
<td>-3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>14</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>-1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>-3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>-2</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>-2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>-1</td>
<td>-3</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 24: Experiment 3  
(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers. 1</th>
<th>Vers. 2</th>
<th>Vers. 3</th>
<th>Vers. 4</th>
<th>Vers. 5</th>
<th>Vers. 6</th>
<th>Vers. 7</th>
<th>Vers. 8</th>
<th>Vers. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>-3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>3</td>
<td>-3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Row No.</td>
<td>Vers.1</td>
<td>Vers.2</td>
<td>Vers.3</td>
<td>Vers.4</td>
<td>Vers.5</td>
<td>Vers.6</td>
<td>Vers.7</td>
<td>Vers.8</td>
<td>Vers.9</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>-3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>-2</td>
<td>2</td>
<td>0</td>
<td>-3</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>-3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-3</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>-1</td>
<td>3</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
<td>-3</td>
<td>-3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 27: Experiment 3  

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Vers.4</th>
<th>Vers.5</th>
<th>Vers.6</th>
<th>Vers.7</th>
<th>Vers.8</th>
<th>Vers.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 27: Experiment 3

f) Friendly-Unfriendly

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Vers.1</th>
<th>Vers.2</th>
<th>Vers.3</th>
<th>Vers.4</th>
<th>Vers.5</th>
<th>Vers.6</th>
<th>Vers.7</th>
<th>Vers.8</th>
<th>Vers.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>-2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

(Attitude scores - each Version/Row No. by a different subject)
7. AN EXPERIMENT TO COMPARE RANDOM VERSUS NON-RANDOM PATTERNS OF L.S.E. IN A COMPUTER INTERVIEW

7.1 Aims

Experiments 1 and 2 have indicated that random Encouragement (E.P.s) and Chattiness (C.P.s) have a significant effect on people's attitude towards computer interviews.

Yet, a fourth experiment was conducted aiming to compare the random patterns of L.S.E. with planned patterns of them, as far as the effect of those patterns on interviewees' acceptance of computer interviews is concerned.

This was considered especially important because the randomness of L.S.E. was a rough indication rather than a firm conclusion in the analysis of 4 interviewing programs, the main reason for the latter being the small amount of data (only 4 interview questionnaires being available for the linguistic analysis).

7.2 Material

The questionnaire about leisure time and people's hobbies, mentioned in the previous experiments, was used for this experiment too.

7.3 Variables

The controlled variable of this experiment was the pattern of L.S.E.

This variable had four levels in the experiment, producing four versions of the interview questionnaire. The first level was a random one, with sparse E.P.s and sparse C.P.s (sparse E.P.s and C.P.s as defined in experiments 1 and 2). The second level was a periodic pattern, with same density of E.P.s and C.P.s as in the first level. The third level had the same density of L.S.E. as the two previous levels, but there were more E.P.s and C.P.s towards the end of the interview, the corresponding phrases increasing progressively towards the end of the questionnaire.
The fourth and last level of the controlled variable was decided by two professional interviewers by the following process: First the environment of the computer interview was shown to them, together with information about how the subjects would communicate with the computer. Then, the basic questionnaire with the branching instructions and all the L.S.E. derived from the analysis described in chapter 2 were provided to the interviewers, who were given the chance to ask any questions about the above elements. At the third and last stage the professional interviewers were asked to put the L.S.E. anywhere in the questionnaire, with criterion to maximize the acceptability of the computer interview to the interviewees. (The relevant interview was computer administered, like the interviews with the other three versions. Moreover out of the 20 subjects of this interview half were interviewed by the version of one of the professional interviewers and the other half by the version produced by the second professional interviewer.)

The first level represents the best version (more accurately the most accepted version) of the random pattern of E.P.s and C.P.s, as decided by experiments 1 and 2. The second level was inspired by the graphical representations E.P.s = f (question number) and C.P.s = f (question number) of some of the programs of the analysis, e.g. E.P. = f (question number) for program GASTRIC ULCER. The aim of the third level was to provide extra reinforcement to the interviewee towards the end of the computer interview, where his/her concentration was rather expected to decrease. The fourth level was used in order to compare the intuitive patterns of the people, who wrote the four programs which have been analysed, against the intuitive patterns of people very experienced in face-to-face interviewing. Moreover the two professional interviewers were encouraged to comment on the adequacy of the L.S.E., elicited from the four interviewing questionnaires, for a computer interview. On this point the impression of both interviewers was that the E.P.s and C.P.s were generally adequate for their purpose.
7.4 Subjects

Students of Loughborough University were the subjects of this experiment.

The selection was done by approaching the students at random, and asking them politely to participate in the experiment. Most of the students approached (about 70%) were willing to participate.

A total of 80 students took part in the experiment, 20 per each of the four versions of the questionnaire, with equal number of men and women participating in each version.

7.5 The Computer System

The computer system used during this experiment was exactly the same as that used in experiment 1. These programs were written in an extended version of the BASIC computer language running, for this experiment, in the PDP 11/10 mini-computer of the Department of Human Sciences of Loughborough University.

Exactly the same with that of the first experiment was also the environment during this interview, and so was the way the user could communicate with the system (i.e. the questions were again presented on a V.D.U. similar to those of experiments 1 and 2). The only difference was the speed of presenting the questions, which was 10 characters/second for this experiment. Moreover, unlike experiment 1, the computer system conducted the main interview only and not the attitude measurements after the interview.

7.6 The Man-Computer Interaction

The interaction during the computer interview followed exactly the same steps as in experiment 1.

Each person of the 80 interviewees had only one interview with the computer on one of the four questionnaire versions, the interviewees being randomly assigned to one of the four versions.
As in experiment 1, each subject was told that the computer was going to ask him/her a few simple YES/NO questions about leisure time and hobbies. He/she was also told that he/she could answer the questions, which appeared on a V.D.U., by pressing one of the three buttons of a keyboard labelled 'YES', 'NO', or '?', while the computer would assist him/her throughout the whole process.

Before leaving the room, the experimenter initialised the system, and then the subject was left alone in front of a keyboard and a V.D.U. During the interaction, the user was communicating with the system by pressing the three buttons of the keyboard, while the system presented the questions in natural English on a V.D.U. at a rate of 10 characters/sec.

The interview went on smoothly and no problem was reported by any of the 80 subjects.

7.7 The Semantic Differential Method

As for experiments 1, 2 and 3, the attitude of people towards the interview was assessed by the Semantic Differential method.

The six bipolar adjectives of factor 1 were used, which were decided by experiment 1. Each of the two poles of each adjective was presented at the end of a part of a line separated in seven equidistant spaces, the scale position labelled with the adverbs slightly, quite, extremely around the neutral position for reasons similar to those given in experiment 1.

An example of the format of presentation of those six adjectives to the people after the computer interview is given in figure 31. These adjectives were presented to the subjects by the experimenter immediately after the computer interview using paper and pencil.
7.8 Analysis of Ratings

The ratings of the concept of the 'computer interview they had' by the 80 interviewees for the six bipolar adjectives are shown in Tables 28-33 correspondingly. To analyse these data the analysis of variance for a single factor experiment (fixed factor) was employed, assuming a linear structural model, as that described in experiment 1.

The result of this analysis was: no significant differences among the four versions at the level of 0.05 and no difference for the combined factor scores too. The factor scores were calculated from the adjectives scores according to the method of the composite estimates (i.e. the scores were first standardised, and then, to calculate the factor scores, the weighted adjective scores were added together, the weight being the square of each adjective's loading on factor 1). The averages and standard deviations of the combined factor scores for the four questionnaire versions are shown in figure 34.

By comparing the first 10 scores of version four (version of interviewer 1), to the subsequent 10 scores of the same version (version of interviewer 2), no significant difference was found between the versions of the two interviewers for all six adjectives and the combined factor score.

The total reaction time and total overall interaction time of each interviewee for all the four versions are given in Tables 34 and 35 correspondingly. The analysis of these data showed no significant results for the reaction time, but presented significant results at $p = 0.05$ for the total interaction time (and these results were further confirmed by analysis of the reaction time per question and the interaction time per question). Further analysis of the latter results showed that this was mainly due to the difference between version 4 and the other three versions combined, while versions 1, 2 and 3 did not differ significantly in the total interaction time.
Table 28: Experiment 4

a) Kind-Cruel
(Attitude scores - each Version/Row No. by a different subject, sex of subjects shown by M or F in table 28 only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 M</td>
<td>2 M</td>
<td>3 M</td>
<td>0 M</td>
</tr>
<tr>
<td>2</td>
<td>2 F</td>
<td>1 M</td>
<td>2 M</td>
<td>3 F</td>
</tr>
<tr>
<td>3</td>
<td>2 M</td>
<td>2 M</td>
<td>2 M</td>
<td>2 M</td>
</tr>
<tr>
<td>4</td>
<td>2 M</td>
<td>2 F</td>
<td>2 F</td>
<td>0 M</td>
</tr>
<tr>
<td>5</td>
<td>2 M</td>
<td>2 F</td>
<td>1 F</td>
<td>2 M</td>
</tr>
<tr>
<td>6</td>
<td>2 F</td>
<td>0 M</td>
<td>0 F</td>
<td>2 F</td>
</tr>
<tr>
<td>7</td>
<td>3 F</td>
<td>2 F</td>
<td>1 M</td>
<td>0 F</td>
</tr>
<tr>
<td>8</td>
<td>0 F</td>
<td>0 M</td>
<td>3 M</td>
<td>2 F</td>
</tr>
<tr>
<td>9</td>
<td>3 F</td>
<td>2 F</td>
<td>0 M</td>
<td>0 M</td>
</tr>
<tr>
<td>10</td>
<td>0 M</td>
<td>2 M</td>
<td>2 F</td>
<td>3 F</td>
</tr>
<tr>
<td>11</td>
<td>3 M</td>
<td>0 M</td>
<td>0 F</td>
<td>3 F</td>
</tr>
<tr>
<td>12</td>
<td>2 M</td>
<td>3 F</td>
<td>2 M</td>
<td>0 M</td>
</tr>
<tr>
<td>13</td>
<td>2 F</td>
<td>0 M</td>
<td>3 F</td>
<td>0 F</td>
</tr>
<tr>
<td>14</td>
<td>3 M</td>
<td>0 M</td>
<td>2 M</td>
<td>3 F</td>
</tr>
<tr>
<td>15</td>
<td>2 M</td>
<td>0 F</td>
<td>0 M</td>
<td>3 M</td>
</tr>
<tr>
<td>16</td>
<td>0 F</td>
<td>3 M</td>
<td>0 F</td>
<td>1 M</td>
</tr>
<tr>
<td>17</td>
<td>3 F</td>
<td>0 F</td>
<td>0 M</td>
<td>0 F</td>
</tr>
<tr>
<td>18</td>
<td>0 M</td>
<td>0 F</td>
<td>2 F</td>
<td>3 F</td>
</tr>
<tr>
<td>19</td>
<td>0 F</td>
<td>3 F</td>
<td>0 F</td>
<td>0 M</td>
</tr>
<tr>
<td>20</td>
<td>2 F</td>
<td>3 M</td>
<td>2 F</td>
<td>0 F</td>
</tr>
</tbody>
</table>
Table 29: Experiment 4  

b) Graceful-Awkward  
(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>18</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>-1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 30: Experiment 4

c) Pleasant-Unpleasant

(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 31: Experiment 4

d) Interesting-Boring

(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 32: Experiment 4

e) Relaxed-Tense
(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>-1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 33: Experiment 4

f) Friendly-Unfriendly
(Attitude scores - each Version/Row No. by a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 34: Experiment 4: Reaction Time (in seconds)  
(Each Version/Row No. refers to a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>166.8</td>
<td>301.8</td>
<td>220.8</td>
<td>110.4</td>
</tr>
<tr>
<td>2</td>
<td>336.0</td>
<td>163.8</td>
<td>164.4</td>
<td>382.2</td>
</tr>
<tr>
<td>3</td>
<td>288.0</td>
<td>264.0</td>
<td>245.4</td>
<td>141.6</td>
</tr>
<tr>
<td>4</td>
<td>480.0</td>
<td>305.4</td>
<td>345.6</td>
<td>223.2</td>
</tr>
<tr>
<td>5</td>
<td>205.2</td>
<td>332.4</td>
<td>336.6</td>
<td>197.4</td>
</tr>
<tr>
<td>6</td>
<td>304.2</td>
<td>514.8</td>
<td>153.0</td>
<td>202.2</td>
</tr>
<tr>
<td>7</td>
<td>87.6</td>
<td>265.8</td>
<td>213.6</td>
<td>288.0</td>
</tr>
<tr>
<td>8</td>
<td>290.4</td>
<td>154.8</td>
<td>202.2</td>
<td>244.8</td>
</tr>
<tr>
<td>9</td>
<td>218.4</td>
<td>163.8</td>
<td>222.0</td>
<td>235.8</td>
</tr>
<tr>
<td>10</td>
<td>153.0</td>
<td>213.0</td>
<td>187.8</td>
<td>145.8</td>
</tr>
<tr>
<td>11</td>
<td>205.2</td>
<td>284.4</td>
<td>126.6</td>
<td>189.0</td>
</tr>
<tr>
<td>12</td>
<td>234.0</td>
<td>300.6</td>
<td>356.4</td>
<td>633.4</td>
</tr>
<tr>
<td>13</td>
<td>187.2</td>
<td>252.6</td>
<td>421.2</td>
<td>245.4</td>
</tr>
<tr>
<td>14</td>
<td>258.6</td>
<td>394.2</td>
<td>378.0</td>
<td>564.0</td>
</tr>
<tr>
<td>15</td>
<td>310.8</td>
<td>144.0</td>
<td>271.2</td>
<td>228.0</td>
</tr>
<tr>
<td>16</td>
<td>186.0</td>
<td>589.2</td>
<td>189.6</td>
<td>412.8</td>
</tr>
<tr>
<td>17</td>
<td>135.6</td>
<td>210.6</td>
<td>314.4</td>
<td>160.8</td>
</tr>
<tr>
<td>18</td>
<td>350.4</td>
<td>263.4</td>
<td>259.2</td>
<td>269.4</td>
</tr>
<tr>
<td>19</td>
<td>148.2</td>
<td>195.6</td>
<td>265.2</td>
<td>299.4</td>
</tr>
<tr>
<td>20</td>
<td>255.6</td>
<td>206.4</td>
<td>226.2</td>
<td>229.2</td>
</tr>
</tbody>
</table>

(per subject) $\bar{X} = 4801.2$  
S.D. = 91

(per subject and per question) $\bar{X} = 8.2$  
S.D. = 3.9

Total number of questions = 28 with 3 introductory questions.
Table 35: Experiment 4: Total Interaction Time (seconds)

(Each Version/Row No. refers to a different subject)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1082.4</td>
<td>1001.4</td>
<td>879.0</td>
<td>969.0</td>
</tr>
<tr>
<td>2</td>
<td>1008.0</td>
<td>988.8</td>
<td>829.2</td>
<td>1095.6</td>
</tr>
<tr>
<td>3</td>
<td>1323.0</td>
<td>991.8</td>
<td>1032.6</td>
<td>849.6</td>
</tr>
<tr>
<td>4</td>
<td>1153.8</td>
<td>1166.4</td>
<td>952.8</td>
<td>1165.8</td>
</tr>
<tr>
<td>5</td>
<td>868.2</td>
<td>1180.8</td>
<td>1174.2</td>
<td>885.0</td>
</tr>
<tr>
<td>6</td>
<td>1062.8</td>
<td>969.6</td>
<td>1006.8</td>
<td>894.0</td>
</tr>
<tr>
<td>7</td>
<td>759.0</td>
<td>829.2</td>
<td>877.8</td>
<td>1018.8</td>
</tr>
<tr>
<td>8</td>
<td>969.0</td>
<td>845.4</td>
<td>877.8</td>
<td>925.8</td>
</tr>
<tr>
<td>9</td>
<td>896.4</td>
<td>889.2</td>
<td>1183.8</td>
<td>915.6</td>
</tr>
<tr>
<td>10</td>
<td>1153.8</td>
<td>1166.4</td>
<td>952.8</td>
<td>1165.8</td>
</tr>
<tr>
<td>11</td>
<td>868.2</td>
<td>1180.8</td>
<td>1174.2</td>
<td>885.0</td>
</tr>
<tr>
<td>12</td>
<td>1062.8</td>
<td>969.6</td>
<td>1006.8</td>
<td>894.0</td>
</tr>
<tr>
<td>13</td>
<td>759.0</td>
<td>829.2</td>
<td>877.8</td>
<td>1018.8</td>
</tr>
<tr>
<td>14</td>
<td>969.0</td>
<td>845.4</td>
<td>877.8</td>
<td>925.8</td>
</tr>
<tr>
<td>15</td>
<td>896.4</td>
<td>889.2</td>
<td>1183.8</td>
<td>915.6</td>
</tr>
<tr>
<td>16</td>
<td>1153.8</td>
<td>1166.4</td>
<td>952.8</td>
<td>1165.8</td>
</tr>
<tr>
<td>17</td>
<td>868.2</td>
<td>1180.8</td>
<td>1174.2</td>
<td>885.0</td>
</tr>
<tr>
<td>18</td>
<td>1062.8</td>
<td>969.6</td>
<td>1006.8</td>
<td>894.0</td>
</tr>
<tr>
<td>19</td>
<td>759.0</td>
<td>829.2</td>
<td>877.8</td>
<td>1018.8</td>
</tr>
<tr>
<td>20</td>
<td>969.0</td>
<td>845.4</td>
<td>877.8</td>
<td>925.8</td>
</tr>
</tbody>
</table>

Σ = 20087.8  20137.8  19865.4  22382.4

X = 1004.3  1006.9  993.3  1119.1

S.D. = 154.8  167.6  150.4  216.3

(per subject X = 40.7  44.7  43.7  49.9
and per question) S.D. = 10.7  11.3  8.9  12.3
In parallel to the above results, the versions of the two professional interviewers did not present significant differences between each other for the total reaction time. Yet, they presented significant results for the overall interaction time at the 0.05 level of significance.

No significant differences were also found for the reaction times of the 25 main questions of the questionnaire. (The $\bar{x}$s and S.D.s of the 25 questions are shown in table 34.)

The sex of each interviewee for all bipolar adjectives, and the two times (reaction and overall interaction) are indicated in table 28 by an M for males and an F for females. By analysing these scores for differences between the two sexes no significant difference was found, either for the six adjectives and the combined factor scores, or for the reaction time and the total interaction time.

7.9 Conclusions

This experiment has shown that during a computer interview the random pattern of L.S.E. of optimum density may be as acceptable to the interviewees as more planned patterns of the same density (i.e. a periodic one and one with more E.P.s and C.P.s towards the end of the interview). Moreover it has shown that this random pattern may be as acceptable as those patterns which were devised by professional interviewers aiming to maximise the acceptability of the particular interview to the interviewees.
8. AN EXPERIMENT TO COMPARc PEOPLE'S ATTITUDE TOWARDS COMPUTER AND HUMAN INTERVIEWS

8.1 Aims

The four experiments described so far have investigated whether random E.P.s and C.P.s affect a computer or a human interview, what is the density of the most acceptable version of random E.P.s and C.P.s, and how random patterns of E.P.s and C.P.s are compared with more planned patterns - all of these in relation to the concept of the acceptability of the interviews to the interviewees. As a result of those experiments, an optimum random pattern of E.P.s and C.P.s was found for computer interviews, which was at least as acceptable as more planned patterns (as described in section 7.8).

Yet, it was felt that the study of the usefulness of L.S.E. in a computer interview would not be completed, unless an experiment was conducted to study how this optimum random pattern of E.P.s and C.P.s is compared with human interviews, as far as the acceptability of the two methods to the interviewees is concerned. The latter was the aim of the fifth experiment described below.

8.2 Material

The same basic questionnaire about leisure time and people's hobbies, used in the previous four experiments, was used for this experiment too.

8.3 Variables

The controlled variable of this experiment was the method of interviewing. Two levels of this variable were used. The first one was a computer interview with 'sparse' random Encouragement and Chattiness. The second level consisted of human interviews conducted by two professional interviewers.
8.4 **Subjects**

The subjects of this experiment were students of Loughborough University.

The potential subjects were approached at random in the campus of the University, and were asked politely to participate in the experiment. Then the first twenty, of those who gave a positive response (refusals about 30%), were chosen to take part in the experiment.

Each subject was interviewed by the computer and one of the two interviewers. Moreover, due to a balancing process, the same number of men and women participated in this experiment and were balanced equally between the two interviewers (who were both female).

8.5 **The Computer System**

The computer system used in this experiment was the same as the system used in experiment 4 (described in section 7.5), written in an extended version of the BASIC computer language and running in a PDP 11/10 mini-computer.

8.6 **Man-Computer Interaction versus Man-Man Interaction**

The ten men and ten women, who took part in the experiment, had one interview with the computer and one with one of the two interviewers each, the order of the method of the interview being balanced.

During the computer interview the questions were presented by the system in natural English on a V.D.U. at a rate of 10 characters/sec., while the interviewees could answer the questions by pressing one of the three buttons of a keyboard labelled 'YES', 'NO', and '?'.

Each subject had two interviews, one approximately 10 minutes after the other.
During the human interview the experienced interviewers were absolutely free to ask the questions in any format and style they liked, but following basically the same sequence of questions as the computer.

All the subjects reported no problems for either of the two methods.

8.7 The Semantic Differential Method

To assess the attitude of the subjects, after the computer and human interviews, the Semantic Differential method was employed, exactly as with the previous four experiments.

Again the six related bipolar adjectives, which constitute the factor Personal Evaluation, were used after each interview (computer and human) in a way similar to that of the previous experiments.

8.8 Analysis of Results

The scores of the 20 subjects on each bipolar adjective for the computer and human interview are shown in tables 36 and 37. The sex of the subjects is indicated in column 2 of table 36 and with an M for males and an F for females. The first 10 subjects had an interview with the first professional interviewer, while the last 10 with the second one (all of the subjects had an interview with the computer). The averages and standard deviations of the scores of the subjects, based on tables 36 and 37, are given in table 38.

To analyse the data of tables 36 and 37 and those of the combined factors scores (calculated from the adjectives scores by the method of composite estimates), the analysis of variance for a single factor experiment with repeated measures was employed.
Table 36: Experiment 5: The first three Adjectives of Factor 1
(Attitude rating scores – same subject across each row)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>-1</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>-1</td>
<td>8</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>-1</td>
<td>10</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>3</td>
<td>2</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>12</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>3</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>0</td>
<td>3</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>3</td>
<td>2</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>2</td>
<td>1</td>
<td>19</td>
<td>-1</td>
<td>0</td>
<td>19</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>-1</td>
<td>20</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 37: Experiment 5: The last three Adjectives of Factor 1
(Attitude rating scores - same subject across each row)

d) Interesting-Boring
e) Relaxed-Tense
f) Friendly-Unfriendly

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>-1</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>3</td>
<td>7</td>
<td>-1</td>
<td>2</td>
<td>7</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>-1</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>-1</td>
<td>2</td>
<td>11</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>3</td>
<td>3</td>
<td>13</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>3</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>14</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>-1</td>
<td>1</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>3</td>
<td>18</td>
<td>3</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>0</td>
<td>19</td>
<td>-1</td>
<td>1</td>
<td>19</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>3</td>
<td>-1</td>
</tr>
</tbody>
</table>
Table 38: Experiment 5: The Scores on each Bipolar Adjective of Factor 1

<table>
<thead>
<tr>
<th>Bipolar Adjective</th>
<th>Average Human</th>
<th>Average Computer</th>
<th>Standard Deviation Human</th>
<th>Standard Deviation Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kind-Cruel</td>
<td>1.55</td>
<td>1.35</td>
<td>1.28</td>
<td>1.09</td>
</tr>
<tr>
<td>Graceful-Awkward</td>
<td>0.50</td>
<td>0.55</td>
<td>1.24</td>
<td>1.28</td>
</tr>
<tr>
<td>Pleasant-Unpleasant</td>
<td>1.60</td>
<td>1.15</td>
<td>1.31</td>
<td>1.09</td>
</tr>
<tr>
<td>Interesting-Boring</td>
<td>1.00</td>
<td>1.80</td>
<td>1.34</td>
<td>1.06</td>
</tr>
<tr>
<td>Relaxed-Tense</td>
<td>1.55</td>
<td>1.75</td>
<td>1.39</td>
<td>1.16</td>
</tr>
<tr>
<td>Friendly-Unfriendly</td>
<td>2.15</td>
<td>1.15</td>
<td>1.23</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Two of the six bipolar adjectives showed significance of the data at the 0.05 level, namely the adjectives Interesting-Boring and Friendly-Unfriendly. Out of those two bipolar adjectives, the first showed an attitude of the people favourable to the computer, while the second one showed an attitude favouring the human interviews. The other four bipolar adjectives showed no significance of the data and so did the combined factor scores, the latter presenting almost equal averages for the two methods.

Comparing the data obtained by the two professional interviewers, no significant difference between them was found for the six bipolar adjectives and the combined factor scores.

Analysing the data for sex differences, significant results were found at the 0.05 level for the adjective Kind-Cruel and for the computer interview only, with females scoring the computer interview as significantly more 'kind' than the male subjects.
8.9 Conclusions

This experiment has shown that computer interviews (if using random Encouragement and Chattiness of optimum density), may be as acceptable to the interviewees as human interviews.

This is the last of a series of five experiments to measure the effect of context-free L.S.E. on people's attitude towards computer and human interviews. These five experiments have demonstrated that a computer interview using a questionnaire 'dressed up' with random E.P.s and C.P.s (of optimum density) may be: a) as acceptable as that using more planned patterns of E.P.s and C.P.s, and b) as acceptable as human interviews.
9. COMBINATION AND REVIEW OF RESULTS FROM THE FIVE EXPERIMENTS AND GENERAL DISCUSSION THEREON

9.1 Summary of Thesis So Far

An analysis of four highly successful medical history-taking computer interviewing programs (described in Chapter 2) has revealed the L.S.E., i.e. phrases added to the questions and types of questions differing from the S.S.Q.s, which appear to aim at increasing the variability of the interaction and therefore the interviewee's acceptance of it. Moreover the same analysis has found that: a) the E.P.s and C.P.s (representing the L.S.E.) follow a rather random pattern among the questions, and b) about 70% of the L.S.E. were context-free i.e. independent of the meaning of the underlying S.S.Q.s.

The five experiments following the analysis, which were described in chapters 4-8, were aiming at measuring the effect of the L.S.E., more precisely of the context-free L.S.E., on people's attitude towards computer and human interviews. The general plan was as follows.

First, the effect of random Encouragement and Chattiness (the pattern derived from the previous analysis) on people's acceptance of computer interviews was measured. Accordingly one experiment (described in chapter 4) was devoted to measuring the effect of both the E.P.s and C.P.s combined, while a second experiment (described in chapter 5) was conducted to measure the effect of the variable Encouragement only. These two experiments have shown that: a) random Encouragement and Chattiness may have a significant effect on people's acceptance of computer interviews, and b) there appears to be an optimum in the amount of random E.P.s and C.P.s used in a computer interview.
Second, having found that the random patterns of E.P.s and C.P.s are 'useful' for computer interviews, it was thought to be interesting to examine the effect of these patterns on interviewee's acceptance of human interviews. The relevant experiment (described in chapter 6) has shown that, in contrast to computer interviews, random E.P.s and C.P.s have no effect on people's acceptance of human interviews, while random E.P.s might even have a negative effect.

Third, having measured the effect of random Encouragement and Chattiness on people's attitude towards computer (and human) interviews, it was thought to be essential to compare the random pattern of E.P.s and C.P.s with more planned patterns of these two variables, as far as their effect on people's acceptance of computer interviews is concerned. Therefore an experiment was conducted for this purpose (described in chapter 7). This experiment has shown that the random pattern of E.P.s and C.P.s of optimum 'density' may be as 'good' as the planned patterns with which it was compared.

Fourth, after the above results, it was felt that one more experiment would be necessary to complete the study on the effect of L.S.E. on computer interviews, i.e. an experiment to compare computer interviews ('dressed up' with random E.P.s and C.P.s of optimum density) with human interviews. Therefore one more experiment was conducted (described in chapter 8), but has shown no difference between the two examined methods.

The above results are discussed in more detail below.
9.2 The Method for Measurement of Acceptability of Interviews

First of all, during experiment 1, described in Chapter 4, a simple and practical method to measure people's attitude towards the interview has been devised in the form of six related bipolar adjectives (shown in Table 5, factor 1). This device has been used to assess the attitude of the interviewees towards the computer and human interviews in all five experiments described in chapters 4-8.

Moreover the precision of this device in experiment 1 was not different than that of experiments 2, 3, 4 and 5. (A difference of this kind could occur, because in experiment 1 the 6 bipolar adjectives were presented with 14 more bipolar adjectives, while for the remaining four experiments only these 6 bipolar adjectives were presented.) This was shown by comparing the attitude scores on these 6 bipolar adjectives of experiments 1 and 2 (for the same questionnaire version, i.e. Version 1), which resulted in no significant difference between the two groups of scores. The above result also indicates that there was no significant interaction effects between the 6 bipolar adjectives of factor 1 and the remaining 14 bipolar adjectives of factors 2-6 (the 20 bipolar adjectives and the 6 factors as described in Chapter 4).

9.3 The Influence of Encouraging and Chatty Phrases

Then, in both experiments 1 and 2 we have seen that random patterns of L.S.E. (represented by the two variables Encouragement and Chattiness) seem to have a significant effect on interviewee's acceptance of computer interviews \((p < 0.05\) for experiment 1, \(p < 0.01\) for experiment 2).

Moreover, the same two experiments have also indicated that there appears to be an optimum for the amount of random Encouragement and Chattiness used in a computer interview. The optimum seems to be a sparse amount of both (at a rate of about 1 E.P. or C.P. per 3 questions), and neither too few nor too many both of which levels receive lower acceptability ratings. It is important to note that
these results were obtained using a between-subjects experimental design, i.e. any one subject responded to one questionnaire version only.

This optimum amount of Encouraging and Chatty Phrases during the computer interview resembles a man-man interaction situation, where a person uses that format probably by intuition, and was derived from the frequency rate of such phrases found in the initial analysis.

However it is important to note that this optimum, or optimum range, of E.P.s and C.P.s increases if these phrases are added not randomly in an interview questionnaire, but follow a 'logical sequence among the questions'. This was shown by experiment 4, where although one of the two professional interviewers produced a version with significantly more E.P.s and C.P.s than the 'random optimum', this did not result in a less acceptable rating of this version by the people.

9.4 Comparison of Different Patterns of Encouraging and Chatty Phrases

The aim of experiment 4 was to test the adequacy of random patterns of L.S.E. by comparing them with more planned patterns. Accordingly four versions of the same basic questionnaire were compared, the first with random E.P.s and C.P.s, the second with a periodic pattern of the two variables, the third with more E.P.s and C.P.s towards the end of the interview (where the concentration of the interviewees is rather expected to decrease), and the fourth with a pattern and density of the E.P.s and C.P.s as decided by professional interviewers. This last version was particularly important, because it would enable the experimenter to compare the intuitive patterns of the N.P.L. group for the L.S.E. in a computer interview, as reconstructed by the computer, to the patterns of the L.S.E. as decided by experienced interviewers, also presented via the computer, aiming to maximize the acceptability of the particular interview.
However the four versions produced no significant results on the acceptance of people towards the interview. This result indicates that random Encouragement and Chattiness of a certain density can be as acceptable as more planned patterns of the two variables (the ones used in this experiment), and as acceptable as the patterns of E.P.s and C.P.s decided by experienced interviewers. By using the same number of male and female interviewees for each of the four versions of the questionnaire, an analysis of the data for sex differences was attempted. This resulted in no significant results, indicating roughly the same degree of acceptance of computer interviews for the two sexes. Moreover, it was found that the two professional interviewers (both females) produced versions of about the same acceptance to the interviewees.

Therefore experiment 4 has further reinforced the usefulness of the random patterns of L.S.E. in a computer interview, by showing that they may be as acceptable to the interviewees as more planned patterns. Yet, one more test had to be done to complete the study on the usefulness and adequacy of random context-free L.S.E. of optimum density in a computer interview; i.e. to compare a computer interview, its questions being 'dressed up' with random context-free L.S.E. of optimum density, with human interviews as far as their acceptance to the interviewees is concerned. Experiment 5 was conducted to settle the issue. The relevant discussion on this subject is present in section 9.7.

9.5 The Reaction Time and Interaction Time Data

For both experiments 1 and 2 the average reaction time (both per subject, and per subject and per question) i.e. the time it took the subjects to press a 'YES', 'NO' or '?' button after the presentation of each question, was longer for version 2 (optimum amount of E.P.s and C.P.s). It is possible that this difference is due to the fact that the subjects took more seriously the most acceptable version 2, and spent more time thinking about the questions, while for the less acceptable versions 1 and 3 they tried to finish quicker. However, the relatively small number of subjects
(18-20 for each of the three versions) did not permit a statistical confirmation of these results, and it would be interesting to see experiments of this kind, but with more subjects, conducted to clear this point.

The total interaction time (both per subject and per subject and per question), i.e. time required for presenting the questions and giving the answers, presented significant results for experiment 1 \((p<0.05)\), with version 3 having a higher average than version 2 and version 2 than 1. (Version 1 with no E.P.s and C.P.s, version 2 with sparse amount of the two variables and version 3 with dense amount of the variables - sparse and dense variables as defined in chapter 4). However this significant result was very probably due to the fact that version 3 has considerably more total number of words in it, than both the other versions, and that version 2 had more words than version 1, in relation to the relatively low speed of presentation of the questions (30 characters/sec.). This fact was further supported by the results of experiment 2, where, because of the higher speed of presenting the questions (>100 characters/sec.), the three versions presented no significant differences of the total interaction time.

The total interaction time (both per subject and per question and subject) presented significant results at the level of 0.05 for experiment 4 too. This result was, as further statistical analysis has shown, due to the difference of version 4 (devised by two professional interviews) and the other three versions combined, while versions 1, 2 and 3 did not differ significantly (versions 1, 2 and 3 all with sparse amount of E.P. and C.P.s but with different patterns of these two variables). Moreover, there was a significant difference at the 0.05 level of the total interaction time between the two versions revised by the two professional interviewers (the two versions of the professional interviewers form the experimental version 4). The above two significant results of the total interaction time were clearly due to the longer questions and more phrases of the version produced by one of the two
professional interviewers as shown by observing the length of the questions, in combination with the slow speed of presenting the questions (10 characters/sec.). The fact that one of the two professional interviewers produced a version with more E.P.s and C.P.s than the other interviewer and the N.P.L. group possibly reflects the difference of attitude of the above people towards 'computers' and 'the capabilities of computers to carry out interviews'. Accordingly one of the two professional interviewers probably had a more 'negative' attitude towards the computer and its capabilities for carrying out interviews than the other interviewer and the N.P.L. people, and therefore she produced a version with more E.P.s and C.P.s, using more phrases as a compensation for the computer's 'incapability'.

Moreover, as experiment 4 has shown, no significant differences were found between the reaction times of the 25 main questions of the questionnaire. It is important to note that the lack of differences of reaction times among the questions presumably holds for this kind of 'neutral' questionnaire only, while one would expect significant differences among questions of a more 'sensitive' questionnaire, reflecting the 'emotional' differences, which 'sensitive' questions might cause to the interviewees.

By comparing reaction times per question of the same questionnaire version, but for different speeds of presentation of the questions, no significant results were found. (Table 10 shows reaction times for 30 character/sec. presentation speed, Table 16 for 100 characters/sec. and Table 34 for 10 characters/sec. Versions 1 of both tables 10 and 16 consist of the same questionnaire version, while versions 2 of table 10 and 1 of table 34 also represent the same questionnaire version.) However the 30 characters/sec. data presented a much lower average reaction time than both the 10 characters/sec. and the 100 characters/sec. data. Is this result merely a random variation due to experimental error, or is it a genuine difference, which the between-subjects comparison for this small sample (20 subjects per version) was not sensitive enough to
show? And for the latter case does this result represent an optimum of the speed of presentation in relation to reaction time, in analogy to the optimum amount of random E.P.s and C.P.s, or is it due to the sample difference, since both tables 16 and 34 show data elicited from students, while table 10 (the optimum speed) shows data with subjects from the general public? It would be interesting to see experiments, perhaps using a within-subjects design or alternatively a between-subjects design but with more subjects, to clear these points.

The total interaction times per question of the same questionnaire versions, but for different speed of presenting the questions, did show significant results with lower speeds of presentation showing a significant higher total interaction time. (Table 11 shows total interaction times for 30 characters/sec., table 17 for 100 characters/sec. and table 35 for 10 characters/sec. The same questionnaire versions as for the reaction times above were compared. The results were significant at the 0.05 level, and for the comparison with the data of table 35 at the 0.001 level). These significant results were clearly due to the different times required by the questions for different presentation speeds to appear on the screen, and were to be expected. However one question which comes into mind, following the above significant results on the total interaction time, is whether the different speeds of presentation have any effect on the attitude scores too.

9.6 Comparison of Attitude Scores

Yet, by comparing the attitude scores for the above questionnaire versions but for different speeds of presentation, as for the reaction time and total interaction time above, no such significant differences were obtained. (Figures 30, 32 and 34 show the averages and standard deviations of attitude scores for speeds of presentation of 30 characters/sec., 100 characters/sec. and 10 characters/sec. correspondingly.) Moreover there was not much difference between the average attitude scores for the different speeds of presentation.
These results suggest that, for this 'neutral' kind of questionnaire, there is no interaction between the speed of presentation of the questions and the attitude scores of people towards the interview, for speeds of presentation ranging from 10 characters/sec. to 100 characters/sec. (i.e. a practically instantaneous presentation).

9.7 The Computer and Human Interviews Compared

Experiments 1 and 2 have shown that random E.P.s and C.P.s may have a significant effect on interview's acceptability. The fact that the random patterns of Encouraging and Chatty phrases proved acceptable during computer interviews might indicate that the subjects, during the short computer interviews (about 10 minutes), experienced a dialogue simulating in a way a man-man interaction situation, in which they would probably accept random Encouragement and Chattiness during its first exploratory stages, but would possibly require a more planned pattern of those variables as the interaction goes on and the level of knowledge and understanding of the interacting members grows.

In further support of the above, experiment 3 has shown that in a man-man interviewing situation, in contrast to a man-computer situation, random Encouragement and Chattiness do not seem to improve the acceptability of the interview to the interviewee. Moreover random Encouragement was shown to have a rather negative effect on acceptability. (This was also indicated by the spontaneous comments of the professional interviewers who commented that the versions with high Encouragement and Chattiness seemed 'unnatural' or 'unhuman'.) These two apparently different results for computer and human interviews possibly indicate that the interviewees can accept a computer (a machine) interacting like an imperfect human, but cannot accept this from a human.

To enable a direct detailed comparison between computer and human interviews, in relation to their acceptance to the interviewee, a within-subjects experiment was considered (experiment 5) (comparing version 2 of experiment 1 with the same version of experiment 3, i.e. version 5 did not show any significant differences between the two methods in a between-subjects comparison).
within-subjects design was chosen rather than the between-subjects one, because although it is less reliable than the between-subjects design, it is more sensitive than the second, i.e. would detect smaller differences of attitude of the interviewees towards the two methods.

The most accepted version, among those with the E.P.s and C.P.s following a random pattern, was used (as decided by experiments 1 and 2), for the computer interview, while the two professional interviewers, who carried out the human interviews, were absolutely free to use the phrases they wished to, when and where they wished to.

The results did not show any significant differences for the combined factor scores nor for four out of the six related bipolar adjectives. However for two bipolar adjectives the results did present significant differences, one in favour of the computer interview, the other in favour of the human interview. Accordingly the twenty male and twenty female interviewees found the computer interview more interesting than the human one (p > 0.05), while the reverse was the situation as far as the friendliness of the interview is concerned (p > 0.01). The former result, i.e. the computer interviews found to be more interesting than the human interviews, was probably due to the novelty of the situation (computer interview), as most of the interviewees were completely naive users, and also possibly due to the acceptance of the computer interview by the interviewees as a useful and time saving application of computers in real life problems. The latter result, i.e. the human interviews found to be more friendly than the computer interviews might reflect the expected attitudes of the interviewees towards the two methods. However if the attitude scores of the interviewees do not reflect expectancy, but consist of a 'genuine' rating of the two methods, it is possible that this result reflects the fact that the adjective 'friendly' is a word closely related to human beings, and therefore it is not very well comprehended if used for the computer, an inanimate machine.
The data produced by the two professional interviewers, who conducted the human interviews, did not present any significant differences between them either for the six related bipolar adjectives, or for the combined factor scores. Comparing the data of experiment 5 for sex differences, no significant results were also found for all six bipolar adjectives apart from the adjective kind, on which the female subjects rated the computer interview as significantly more kind than the male subjects (p < 0.05). The latter could probably be related to the more emotional character of the female subjects than that of the male subjects, although these results become fuzzy, because of the lack of similar support from the other five bipolar adjectives.

Therefore experiment 5 (within-subjects design) has further confirmed the usefulness of random context-free L.S.E. of optimum density for a computer interview, by showing that such a computer interview may be as acceptable to the interviewees as the human interviews.
10. MEASUREMENTS OF REPRODUCIBILITY AND ACCURACY IN A COMPUTER INTERVIEW

The five experiments on people's attitude towards computer and human interviews, which were described in chapters 4-9, have indicated the following:

a) The use of random Encouragement and Chattiness has a significant effect on interviewee's acceptance of human interviews.

b) There appears to be an optimum for the amount of L.S.E. used in a computer interview, with more L.S.E. or less L.S.E. resulting in less accepted versions.

c) The random patterns of L.S.E. of optimum density are as acceptable to the interviewees as more planned patterns, and as acceptable as the patterns which were devised by professional interviewers aiming to maximize the acceptability of the particular computer interview to the interviewees.

d) The computer interview, when using the optimum questionnaire version, may be as acceptable to the interviewees as the human interview.

However, it is not sufficient that a computer interview should be as acceptable to interviewees as a human interview. An equally important issue is whether the information obtained by a computer interview is equally accurate.

Therefore a study was conducted to answer questions such as: 'What is the degree of reproducibility and accuracy of the information derived during a computer interview, if questionnaire versions differing in their degree of acceptability to the interviewees are used?' (e.g. Do the less accepted versions have the same reproducibility and accuracy as the more accepted one or not?), or, 'Is the computer interview more or less reproducible and accurate than the human interview?'

The terms reproducibility and accuracy of information are used in the context under the following meaning: The reproducibility of the information (data) is a measure of the agreement of the data collected by a particular interviewer, or a particular method of interview, in successive experiments under similar conditions. Therefore the reproducibility of the information is a measure of
the consistency of the interviewer, or the method, and not a measure of bias. On the other hand the accuracy of information measures the deviation between the collected information (data) and the true information, and therefore it measures bias too.

The method for this experiment (described in sections 10.5 and 10.6 below) was to interview a number of subjects by both computer and experienced human interviewers. The answers of the subjects, on a number of independent pieces of information, were then used to estimate reproducibility and accuracy of the information derived during both computer and human interviews. The statistical analysis to estimate accuracy of the information was based on a model by Dawid and Skene (described in section 10.9).

The experiment is described in more detail below.

10.1 Aims

This experiment had two aims:
1) To compare the degree of accuracy and reproducibility of interview questionnaires, differing in their acceptability to the interviewees during a computer interview (i.e. several versions of random E.P.s and C.P.s of different density).
2) To measure the degree of reproducibility and accuracy of the computer interview, and to compare it with that of the human interview.

10.2 Material

The questionnaire used during this experiment was based on the same basic questionnaire about leisure time and hobbies, which was used during the experiments on people's attitude towards computer interviews, slightly changed to produce a more linear structure. This interview questionnaire of YES/NO questions is given in figures 35 (the differences were relatively minor and can be seen by comparing, for example, questions 18 and 20 in both versions in figures 29 and 35).
10.3 Variables

The controlled variable of this experiment was the 'version' of the questionnaire.

This variable had five levels for this experiment. Three levels for the computer interview and the same for the human interviews but modified by the human interviewers. The three versions of the computer interview were exactly the same as those used in experiment 1, described in chapter 4, i.e. one version with no E.P.s and C.P.s; the second version with sparse E.P.s and C.P.s, and the third version with dense E.P.s and C.P.s. The two versions of the human interview were produced by two human interviewers, who, being provided with the basic questionnaire, were given the freedom of changing the wording and the style of the questions, but not their basic meaning, and thus to introduce their personal linguistic style which they thought to suit best to the particular questionnaire.

10.4 Subjects

The subjects used during this experiment were students of Loughborough University, living on the University campus, and were selected exactly as for experiments 4 and 5, described in sections 7.4 and 8.4 correspondingly. About 75% of the people approached were willing to be subjects. In all 36 students took part in the experiment, 18 males and 18 females.

10.5 The Computer System

The computer system was the same as that used in experiments 4 and 5, written in an extension of the BASIC computer language, and running in a PDP 11/10 mini-computer. This computer system is described in section 7.5 above.

10.6 Man-Computer Interaction versus Man-Man Interaction

Each of the 18 men and 18 women, who took part in the experiment, had one interview with the computer and one with each of the two human interviewers who were used in this study.
Moreover the order of the interview was balanced for all six males and six females who were interviewed by the same version of the computer interview (and by the two human interviewers), and this was done for all three groups of 12 subjects who took part in the experiment. Therefore the whole experiment can be seen as comprised of three parts, with 12 interviewees participating in each part, and the order of the interviews being balanced over these 12 interviewees in such a way as to eliminate the interviewers' order effect.

During the computer interview the questions were presented by the system in natural English on a V.D.U. at a rate of 10 characters/sec., while the interviewees could answer the questions by pressing one of the three buttons of a keyboard labelled 'YES', 'NO' and '?'.

During the human interview, the human interviewers were absolutely free to ask the questions in any format and style they liked (but using the same basic sequence of the 25 questions).

Each interviewee, coming to the Human Sciences Laboratories for the experiment, was told that he/she was going to answer a few YES/NO questions about leisure time and hobbies. The subject was also told that he/she was going to be asked the same basic questions by the computer and by two human interviewers, in order to compare the two methods. Moreover each interviewee was urged not to try to repeat his/her answers during the three interviews, but to answer independently, i.e. as he/she was feeling at the time of each interview. This emphasis was repeated after the first and second interview.

Therefore a typical subject would have an interview say with the computer, a break of about ten minutes, and then two interviews with the two human interviewers with a ten minutes break between them, the whole session lasting about one hour. During the three interviews (the order of which was balanced) the same basic questionnaire was used, but changes in the style of the questions could occur, since the human interviewers were free to phrase the questions as they liked. Moreover the sequence of the questions
during the three interviews could change for every different answer, which the subject might give to a branching question.

Before the computer interview each subject was also told that he/she had nothing to worry about this particular interview, because the computer would assist him/her throughout the whole process.

All the subjects reported no problems for either the computer interview or the two human interviews.

10.7 Analysis of Results

Out of the 25 basic questions of the questionnaire about leisure time and hobbies, given in figure 33, 18 were aiming to elicit independent pieces (bytes) of information (corresponding to the indicants in a medical interview), while the remaining 7 were asking questions dependent on the previous 18 questions. Moreover, the 7 'dependent' questions could be skipped for certain answers of the interviewee to the 18 independent ones. These 18 independent questions are the numbers 1, 2, 3, 5, 7, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23 of the questionnaire, shown in figures 33.

To assess the reproducibility and accuracy of the information derived by the computer and human interviews, only the answers of the interviewees to these 18 independent questions, which were answered by all interviewees, were considered. The independence of these 18 questions and the fact that the questionnaire structure was linear for them, i.e. they were answered by all the interviewees, provided data satisfying on this point the statistical models used.

The answers of the 36 interviewees to the 18 independent questions for each of the three versions of the computer interview are shown in tables 39, 40 and 41 correspondingly. For these data the positive responses are indicated by a 1, and the negative ones by an 0. The three rows of answers of each interviewee to each question correspond to the first human interview, the second human interview, and the computer.
<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Number of Independent Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>4</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>6</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>7</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>10</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>11</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>12</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Subject No.</td>
<td>1</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>111</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
</tr>
<tr>
<td>8</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>110</td>
</tr>
<tr>
<td>10</td>
<td>111</td>
</tr>
<tr>
<td>11</td>
<td>111</td>
</tr>
<tr>
<td>12</td>
<td>110</td>
</tr>
</tbody>
</table>
Table 41: Version 3. Answers of the interviewees to 18 independent questions

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Number of independent question 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 1 0 0 1 1 1 0 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 1 1 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 0 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td>2</td>
<td>1 1 1 1 0 1 1 0 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 0 1 1 0 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 0 1 1 0 0 1 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td>3</td>
<td>1 1 0 1 1 1 1 0 0 1 0 1 0 1 0 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 1 0 0 1 0 1 0 1 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 0 0 0 0 1 0 1 0 1 1 1 1</td>
</tr>
<tr>
<td>4</td>
<td>1 1 0 0 1 1 1 1 1 0 1 0 1 0 1 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1</td>
</tr>
<tr>
<td>5</td>
<td>1 1 0 0 1 1 1 1 0 0 1 1 0 1 1 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>6</td>
<td>1 1 0 1 1 1 0 0 0 1 0 0 1 1 1 1 1 0 0</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 1 1 0 0 1 0 0 1 1 1 1 1 1 0 0</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 1 1 0 0 0 1 0 1 1 1 1 1 1 1 0</td>
</tr>
<tr>
<td>7</td>
<td>1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 0 1 1</td>
</tr>
<tr>
<td>8</td>
<td>1 1 0 1 1 1 0 0 0 0 0 0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 1 1 0 0 0 0 0 0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 0 1 1 1 0 0 0 0 0 0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td>9</td>
<td>1 0 1 0 1 1 1 1 1 1 1 0 0 1 0 0 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 0 1 0 1 1 1 1 1 1 1 0 0 1 0 0 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 0 1 0 1 1 1 1 1 1 1 0 1 0 0 1 0 1 1</td>
</tr>
<tr>
<td>10</td>
<td>1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 0 0 0 0 0 0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 0 1 1 1 1 0 0 0 0 0 0 1 0 1 1 0 1 1</td>
</tr>
<tr>
<td>11</td>
<td>1 1 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>12</td>
<td>1 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1 0 1 1</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1 0 1 1</td>
</tr>
</tbody>
</table>
Furthermore only the answers to the 16 out of the 18 questions were considered in the tests described below, because questions 1 and 11 in the questionnaire of figure 33, which are numbers 1 and 6 in tables 39-41, are introductory questions, and should elicit a positive response for the interview to continue, which exactly was the case for all 36 interviewees.

To assess the reproducibility of the information, the agreements of each interviewer with the consensus were used; the results are given next (in section 10.8). To assess the accuracy of the information a method by Dawid and Skene was used; this method and the results are given in section 10.9. Finally the rationale of this experiment is discussed in chapter 11.

10.8 Reproducibility Results

First, the results of the first interview only, out of the three interviews which each subject had, were compared. These results are given in table 42. (This comparison uses data unaffected by any possible 'memory effect' from previous interviews on the same subject.)

<table>
<thead>
<tr>
<th>Question No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Statistical analysis of the above data has shown that the results of the two human interviewers agree in all 16 independent questions, while the combined results of the two human interviewers and the computer agree in 15 out of the 16 questions. Based on these results the hypothesis that an agreement (or a disagreement) between the results recorded by the human interviewers and the computer is a
random event (null hypothesis) was tested. This hypothesis was rejected in favour to a good agreement between the data recorded by the human interviewers and the computer ($X^2$ test, $p < 0.001$). Therefore, the fact that the frequencies of positive (or negative) answers of the three 12 subject groups to the 16 independent questions have such similarity shows that these three groups belong to the same population of subjects. The above results refer to the first interviews only of all the subjects. The rest of the interviews were then used together with the first ones to assess reproducibility and accuracy of the derived information.

The agreements and disagreements of each interviewer with the consensus were used as an indicator of the reproducibility of the data collected by this interviewer. Based on the data of tables 39, 40 and 41, table 43 shows the disagreements of each interviewer with the consensus for all the three different versions of the computer questionnaire, based on the answers of 36 interviewees to the 16 independent questions. Each of the three columns of results of table 43 (apart from the total column) represents each of the three parts of the experiment, i.e. the data obtained by interviewing 12 subjects (6 males and 6 females) with the interviews order being completely balanced over the 12 subjects to eliminate the interviewers order effect. Moreover the headings version 1, version 2, version 3 of the same table refer to the computer versions only, since the human interviewers were free to use their own style and wording, when asking the questions. (The questionnaire versions 1, 2 or 3 used during the computer interview were assigned at random to each of the 36 subjects.)
Table 43: Disagreements of each interviewer with the consensus

<table>
<thead>
<tr>
<th></th>
<th>VERSION 1</th>
<th>VERSION 2</th>
<th>VERSION 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer 1</td>
<td>8 (4.2%)</td>
<td>6 (3.1%)</td>
<td>13 (6.8%)</td>
<td>27 (4.7%)</td>
</tr>
<tr>
<td>Interviewer 2</td>
<td>4 (2.1%)</td>
<td>5 (2.6%)</td>
<td>4 (2.1%)</td>
<td>13 (2.3%)</td>
</tr>
<tr>
<td>Computer</td>
<td>2 (1.0%)</td>
<td>2 (1.0%)</td>
<td>5 (2.6%)</td>
<td>9 (1.6%)</td>
</tr>
</tbody>
</table>

The percentage in parentheses are the percentages of total disagreement of each interviewer with the consensus, i.e. the 'false positive' disagreements plus the 'false negative' disagreements. (A false positive disagreement indicates that the particular interviewer recorded a 'YES' answer, where the other two interviewers recorded a 'NO' answer. Similarly a false negative disagreement indicates that the interviewer recorded a 'NO' answer, while the other two recorded a 'YES' answer.) The total column shows the number and percentage of each interviewer's disagreements with the consensus, for all the three parts of the experiment, i.e. all the 36 subjects.

Comparing the reproducibility of the three versions of the computer interview (3rd row, without the number of the total column), no significant differences in the results were found ($\chi^2$ test; $\chi^2 = 2$, $p > 0.05$). Moreover no significant differences were found for the reproducibility of every pair of the computer questionnaire ($\chi^2$ test; for version 1 (or 2) and 3 $\chi^2 = 1.3$, $p > 0.05$).

Comparing the reproducibility of the data collected by the two human interviewers and the computer for all 36 subjects (total column), significant differences were found ($\chi^2$ test; $\chi^2 = 10.94$, $p < 0.01$). Further analysis of the data showed that this overall significance was due to the difference of the first human interviewer from the second human interviewer ($\chi^2$ test; $\chi^2 = 4.9$,
p < 0.05), and from the computer (Χ² test; Χ² = 9, p < 0.01),
while the second human interviewer and the computer did not differ
significantly (Χ² test; Χ² = 0.73, p > 0.05).

Before estimating the error rates a comparison of the data
collected from the first interviews only and those collected from the
second and third interviews was attempted to detect possible
distortions of the data collected during the second and third
interviews due to the particular design. The frequency of YES/NO
answers during the 2nd and 3rd interviews are shown in tables 44 and
45 correspondingly. By comparing those data with the data of the
first interviews only (shown in table 42) no significant disagreements
were found (Χ² test; p < 0.01). Moreover the answers of the people
to the second and third interviews were mostly the same with those
of the first interview with very few differences. These differences
were in the expected range for this kind of design as indicated by
similar experiments of other researchers. Therefore no apparent
distortion of the data of the second and third interviews due to the
particular design were detected.

Table 44: Number of positive responses of 36 interviewees
Second interview only

<table>
<thead>
<tr>
<th>Question No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer 1</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>10</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Interviewer 2</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Computer</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 45: Number of positive responses of 36 interviewees
Third interview only

| Question No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|--------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| Interviewer 1| 12| 11| 5 | 8 | 11| 12| 9 | 4 | 1  | 7  | 3  | 5  | 8  | 4  | 9  | 8  | 10 |
| Interviewer 2| 12| 11| 7 | 7 | 11| 12| 11| 6 | 2  | 7  | 4  | 6  | 8  | 5  | 9  | 6  | 10 |
| Computer     | 12| 10| 6 | 7 | 11| 12| 10| 5 | 2  | 7  | 5  | 6  | 8  | 4  | 9  | 5  | 10 |
10.9 Accuracy Results

The estimation of error rates was based on the data of tables 39, 40 and 41 and a statistical model by Dawid and Skene (Dawid and Skene, 1977). This model provides a slow but sure way of obtaining the maximum likelihood estimates of the parameters of interest.

The model makes the following assumptions:
1) For each interviewee the notion of a true but possibly unrecognisable response is sensible.
2) The population of observers interviewing a single interviewee will agree on one response, which, for the purposes of managing this interviewee, can be regarded as the true response.
3) That the above population is consistent from interviewee to interviewee.
4) Most observers interviewing a large number of interviewees will elicit the true response most of the time.
5) The responses given by a single interviewee to successive interviewers are independent. That is the interviewee's reply at any time is not influenced by his/her previous answers.
   Further, all interviewees respond independently.
6) There is no interviewer interviewee interaction. For example, one interviewer does not obtain a more helpful response from some interviewees than is given to other interviewers.

Some of these assumptions in relation to this experiment are discussed in chapter 11.

Before summarising the basic algorithm of the model the following definitions of the symbols used are given: Let us consider K interviewers, indexed as 1, 2, ... K, asking a single question of I interviewees indexed 1, 2, ... I, with each question receiving one of J possible answers. Let \( P_{j}^{(k)} \) be the probability that an observer (K), will record value \( \ell \), when value j is actually the true response and \( M_{i}^{(k)} \) be the number of times interviewer k gets response 1 from interviewee i. Moreover let \( T_{ij} \) (j = 1, 2, ... J; i = 1, 2, ... I) be a set of indicator variables, with \( T_{ij} = 1 \) if \( j = \) the true response for interviewee i, and \( T_{ij} = 0 \) if \( j \neq q \). Further...
let us consider the interviewees I to be a random sample from some population, where the probability that an interviewee drawn at random has true response j is \( p_j \).

The objective of the experiment is to estimate the individual error rates \( p_j^{(k)} \), \( j = 1, 2, \ldots, J; \ell = 1, 2, \ldots, J; k = 1, 2, \ldots, K \). Using the symbols described above, the basic computational algorithm of the model can be summarised as follows:

(i) Take initial estimates of the T's by the equation

\[
T_{ij} = \frac{\sum_k M_{ijk}}{\sum_{i'k} M_{i'jk}} \quad \left( \sum_T = \text{Summation over } T \right)
\]

(ii) Use the following two equations to obtain estimates of the \( p \)'s and \( p'^{\prime} \)'s

\[
p^{(k)}_{j\ell} = \frac{\sum_i T_{ij} \times M_{i\ell k}}{\sum_i \sum_{i'k} T_{ij} \times M_{i'jk}} \quad \text{and} \quad p_j = \frac{\sum_i T_{ij}}{I}
\]

(iii) Use the following equation and the estimates of the \( p \)'s and \( p'^{\prime} \)'s to calculate new estimates of the T's.

\[
T_{ij} = \left( \prod_{k=1}^{K} \prod_{\ell=1}^{J} p^{(k)}_{j\ell} \frac{M_{i\ell k}}{p_j} \right) \frac{\prod_{q=1}^{Q} \prod_{k=1}^{K} p^{(k)}_{q\ell} M_{i\ell k}}{\sum_{q=1}^{Q} \prod_{k=1}^{K} p^{(k)}_{q\ell} M_{i\ell k} p_{\ell q}} \quad \left( \prod = \text{product} \right)
\]

(iv) Repeat steps (ii) and (iii) until the results converge.
For the present experiment $K = 3$, $J = 2$, $I = 12$ for each of the three versions and 36 for the total.

Table 46 summarises the results obtained by applying the algorithm described above to the data of tables 39, 40 and 41.

Table 46: Error rates (a+b) of the interviewers and the computer

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interviewer 1</td>
<td>Interviewer 2</td>
<td>Computer</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>0.070 0.125</td>
<td>0.038 0.092</td>
<td>0.055 0.082</td>
<td>0.115 0.150</td>
</tr>
<tr>
<td></td>
<td>0.125</td>
<td>0.092</td>
<td>0.082</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>0.038 0.092</td>
<td>0.041 0.066</td>
<td>0.091 0.147</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.066</td>
<td>0.135</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>0.055 0.082</td>
<td>0.036 0.049</td>
<td>0.097 0.129</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td>0.049</td>
<td>0.135</td>
<td>0.129</td>
</tr>
</tbody>
</table>

The figures of table 46 show the averages and standard deviations of the false negative (a) errors and the false positive errors (b) taken together, that is of (a+b), for all 16 independent questions of the questionnaire. The estimated errors (a+b) of the three interviewers for each different version of the computer questionnaire have been calculated from the answers of the 12 interviewees who participated in the relevant part of the interview. The total estimated errors (a+b) have been calculated from all three versions together, that is by considering the answers of all 36 subjects who took part in the whole experiment.

By comparing the accuracy of the three different versions of the computer questionnaire no significant results were found (t statistic; \( t_{12} = 0.81, p > 0.05, t_{13} = 1.10, p > 0.05, t_{23} = 1.77, p > 0.05 \)).
By comparing the accuracy of the data collected by the two human interviewers and the computer, no significant results were found too, either for the estimated errors of the computer and each human interviewer (t statistic; \( t_{1-\text{comp.}} = 0.37, p > 0.05, t_{2-\text{comp.}} = 0.04, p > 0.05 \)), or for the estimated errors of the two human interviewers (t statistic; \( t = 0.32, p > 0.05 \)).

10.10 Conclusions

The former results have shown that several versions of the computer interview questionnaire, differing in their degree of acceptability to the interviewees (the acceptability as assessed in the experiment described in chapter 4), elicit information of about the same degree of accuracy. These results agree with the corresponding results on the reproducibility of the information derived by the three versions, which has been found to be roughly the same for the three versions.

The latter results have shown that the computer was as accurate as each of the two human interviewers. Therefore, since the two human interviewers did not differ significantly in the accuracy of the recorded data, we could probably generalise and say that this experiment has shown that the computer may be as accurate as human interviewers. The results on accuracy discussed above differ from those on the reproducibility of the information derived by the two human interviewers and the computer. This difference was due to one of the human interviewers, who was found to have recorded data significantly less reproducible even though apparently no less accurate than those recorded by both the other human interviewer and the computer. Yet, for this case the computer was shown to be (at least) as reproducible as the two human interviewers.
Therefore the following two conclusions can be drawn from this experiment:

a) computer interviewing questionnaire versions, differing in the amount of random E.P.s and C.P.s and in their degree of acceptability to the interviewees, do not differ from each other in the reproducibility and accuracy of the derived information, and

b) computer interviews may be as accurate as human interviews and at least as reproducible as the latter.
DISCUSSING THE RESULTS OF THE EXPERIMENT ON REPRODUCIBILITY AND ACCURACY OF INFORMATION

The statistical analysis on the first interviews only, which was described in the previous chapter, has shown that the 36 subjects, who took part in the experiment, constitute a homogeneous group as far as their answers to the 16 independent questions of the questionnaire are concerned. This is a comfort, since the estimation of reproducibility of the information and the error rates of the interviewers had to be based on the answers of the interviewees to the same basic questions, asked in a different style by the two human interviewers and by the computer within a relatively short period of time. The latter is a constraint of the existing statistical methods and experimental designs, with which all researchers have had to comply in order to measure reproducibility and accuracy of the information (e.g. Evans et al. 1971; Card et al. 1974; Lucas et al. 1977, etc., as discussed in the introductory section 1.1.1).

Yet, the homogeneity of the population together with the uniform conditions of the experiment, i.e. a balanced design to eliminate the interviewer's order effect, seem sufficient in order to permit us to accept the results of this experiment. This is particularly true for the comparison of the results of the three parts of the experiment (involving 12 subjects each), since a distortion of the data on the second and third interviews, should any occur as a result of the particular design, would be expected to be uniform for the three groups. However, no apparent distortion of the data of the second and third interviews due to the particular design was detected.

In the analysis of data of tables 39, 40 and 41 for both reproducibility and accuracy of the derived information, tests were applied to the combined (a+b) disagreements with the consensus and error rates correspondingly. The reason, why no tests were applied to the a's and b's separately, was that the relatively small number of subjects (12 for each of the three different versions of the computer interview questionnaire, i.e. a total of 36) would not provide reliable data for such an analysis. On the other hand,
the combined \((a+b)\)'s averaged over all the 'independent' questions seem reliable enough to be considered and further analysed even for the number of subjects used during this experiment.

For the reproducibility of information, the number of disagreements with the consensus can be considered as a 'good' indicator of it, stable enough even to removal of individual cases.

Therefore the following conclusions can be drawn from the analysis of the data to estimate reproducibility, based on the disagreements of each interviewer with the consensus.

First, the data collected during the computer interview seem highly reproducible. (Maximum disagreement with the consensus 2.6%, for version 3, total disagreement for all versions combined 1.6%.)

Second, it seems that the several versions of the computer questionnaire, differing in the amount of random Encouragement and Chastiness, do not differ significantly in the reproducibility of the derived information.

Third, this experiment suggests that the data collected by a computer interview may be at least as reproducible as those collected by human interviews. It is noticeable that the data collected by the computer were more reproducible than the data collected by both human interviewers and significantly more reproducible than those of one of them. (This same human interviewer produced data significantly less reproducible than those of the other human interviewer too.)

For the accuracy of information, it can be said that in general error rates cannot ever be estimated as there is no way in which the observations recorded by the interviewers can be judged to be sensible or even relevant to the 'true' response, because the 'true' response can never be ascertained.
Yet, error rates can be estimated should the data comply with the assumptions of the model described in section 10.7.

Among those six restrictive conditions only the first part of 5) could cause some concern, while the rest of the assumptions can be considered as tenable for the present case. In particular, the condition of independence of the responses of each interviewee to the three interviewers (two human and the computer) could be violated, because of the short period within which the three successive interviews took place. That is, the interviewees could remember their responses to the same questions, when interviewed by different interviewers, and permit these previous responses to influence their subsequent answers.

To minimise the latter the following two steps were taken during the design of this experiment.

1) During the briefing just before the experiment, and after the first and the second interview, the subjects were urged not to try to give the same response to the questions, should they remember their previous response, but to try to answer the questions 'independently' instead, i.e. to answer the questions as they thought and felt at each time.

2) The interviewer's order effect was eliminated by balancing the subjects (males and females) over the three interviews (two human interviews and the computer interview) for each of the three parts of the experiment, i.e. for each of the three groups of 12 interviewees each.

It could be argued that step 1) was not enough to prevent completely the subjects from being influenced by their previous answers to the same questions, and therefore the absolute values of error estimates can be considered as doubtful.
Yet step 2), the uniform conditions during the three parts of the experiment, and the homogeneity of the 36 interviewees, would certainly permit meaningful comparisons of error rates, both between the versions of the computer interview and between the human interviewers and the computer. For example, if we compare the error rates of one of the human interviewers and the computer, their absolute values could be both slightly different from the real values of these errors, but, due to the uniform conditions and the balanced process, it would be expected that these two errors differ in a similar way from the real values, i.e. are both increased or decreased by say 20% and therefore their comparison is meaningful.

Therefore the following conclusions can be drawn from the analysis of the data of this experiment to estimate accuracy.

First, it seems that the several versions of the computer questionnaire, differing in the amount of random Encouragement and Chattiness, do not differ significantly from each other in the accuracy of the derived information.

Second, this experiment suggests that the data collected by a computer interview may be as accurate as those collected by human interviews. (The computer interview presented the lower average error rate, though the results were not significant. The two human interviewers did not differ significantly too.)
12. **THE CONSTRUCTION OF THE COMPUTER SYSTEM**

One of the aims of this multidisciplinary doctoral research programme was to develop a software system for computer interviewing. The general structure of the system is therefore described in this chapter.

It is unfortunate, yet absolutely true, that different computers (i.e. computers manufactured by different companies) do not have a standard version of a certain computer language, but instead they use different versions which are not compatible one with the other. This makes it costly in time and effort to use a program from one machine on another, even if both machines are supplied with the compiler of the computer language 'common' to both versions. The situation is 'bad' to the extent that some people favour the solution of writing a program from the beginning, rather than modifying it to run on another machine. Even the machines of the same manufacturing company use different versions of the same language, e.g. it was necessary for the programs of this system to be implemented twice, in order to run on a PDP 11/10 and a PDP 11/40 mini-computer.

It is felt that it would be beneficial for users of any kind if standard versions of computer languages were universally accepted, because the application of computers would be spread more, much time and energy would be saved, and the potential of computers to aid the solution of real life problems would be better realised and used. Therefore the author would welcome any organised pressure of the users towards the direction of standardisation of the computer language.

The lack of standardisation of the computer language in use is the main reason why the construction of this computer system is given in the following sections as a set of rules independent of a particular computer language. These rules, since they are independent of a particular language, are general and therefore abstract, but their transformation to a logical flow chart and to specific instructions of a particular computer language would not present many problems for the system analyst and the programmer.
The implementation of this system has been done to the extent described in sections 4.6 and 5.5, because the system was implemented mainly to test the usefulness of the L.S.E., as mentioned in the previous discussion.

12.1 Rules for the Construction of the Computer System

We will restrict the following discussion to YES/NO questionnaires (the possibility of more options is discussed in section 13.5), while a third possibility, '?', has been added for the case when a user does not understand, or is not sure, about his/her answer to a particular question.

The general system can be split into two main parts, which will be discussed below. The first would be able to accept the questionnaire of S.S.Q.s, and the logical structure connecting them, and 'dress them up' to become 'natural' and 'friendly' according to the findings of the analysis (described in section 2). The second would be able to conduct the computer interview keeping trace of the questions asked, the answers given, the reaction times of the users, etc.

The above two parts could have the form of commands, which would be used in order to request from the system the required stage. Moreover, apart from the above two basic commands, other commands could be added to the system, e.g. one command to give a catalogue of the stored questionnaires, one to erase a questionnaire, another to erase or modify a question within a questionnaire, one to list a particular questionnaire, etc. In this way the general system would appear like a command system (e.g. Spiliotopoulos, 1975a).

Before describing the rules for the construction of the basic two parts (stages) of the system, a pre-stage is described, where the rules for the construction of a part of the system, to accommodate the Encouraging and Chatty phrases of the analysis, are given.
12.2 Rules for storing the E.P.s and C.P.s in the System

The E.P.s and C.P.s, as described in section 2, could be stored in a disc file in specified records. By using a vector array \( L_3 (i) \) to specify the length of the \( I \) different E.P.s and C.P.s, numbered 1, 2, \( \ldots \), \( I \), and assuming that the length of each record of the disc file is of constant length (the most usual case) containing say \( Z \) characters, then the number of record and the order of character, where the \( k \)th phrase starts (\( k = 1, 2, \ldots, I \)) can be specified as follows:

\[
\text{Number of Record} = \sum_{i=1}^{k-1} \frac{L_3 (i)}{Z} + 1
\]

\[
\text{Order of Character} = R \left( \sum_{i=1}^{k-1} \frac{L_3 (i)}{Z} \right) + 1
\]

\( R \) = Remainder of

This disc file of E.P.s and C.P.s could be split into five sections corresponding to the five different kinds of phrases, i.e. 'positive' reinforcements, 'negative' reinforcements, rest of E.P.s, pre-phrases of C.P.s, post-phrases of C.P.s. In that case, a five element vector array \( C(5) \) would be needed, to accommodate the cumulative number of the phrases of each of the five sections. Moreover, a vector array \( F(m) \), \( m = I+2 \), would be used, to store the cumulative frequency of the above phrases in the 4 analysed programs of section 2. The last two elements of this array would store the cumulative frequency of the questions with included phrases (Q.I.P.), and the If-Then questions (I.T.Q.) correspondingly.

The above data structures are shown in figure 36. The E.P.s and C.P.s of the four programs, which were analysed, are shown in Appendix 3 together with a typical introduction and thanks at the end of the computer interview.
12.3 Rules for the construction of the first part (stage 1) of the computer system

By initializing the system and using the particular command, the system would be ready to accept a questionnaire and 'dress it up'.

1) Entering the questionnaire. The questionnaire can be entered as a sequence of questions, stored in an alphanumerical one-dimensional array say $A^{\$}(i)$, $i = 1, 2, \ldots N$. Each question could be terminated by a special symbol terminator, while a second terminator could indicate the end of the questionnaire. The length of each question could be stored in an array $L_{1}(i)$, $i = 1, 2, \ldots N$.

After the end of the questionnaire the $N$ triads of numbers, corresponding to the $N$ questions of the questionnaire, would be entered, and stored in three vector arrays say $N_{1}(N)$, $N_{2}(N)$, $N_{3}(N)$. These $N$ triplets indicate the next question after each of the three possible answers of the interviewees to each question.

Then the 'sensitive' questions could be indicated to the system, as $k$ triads of numbers, where the previous analysis has indicated that $k$ is usually around $N/20$, and stored in three vector arrays say $S_{1}(k)$, $S_{2}(k)$, $S_{3}(k)$. The numbers of each of these $k$ triads correspond to the number of each 'sensitive' question in the particular questionnaire, the 'sensitive answer', and the kind of sensitivity of this question-answer. In this case, each of the $S_{1}(i)$, $i = 1, 2, \ldots k$, would take on of the values 1, 2, \ldots $N$. In parallel $S_{2}(i)$, $i = 1, 2, \ldots k$, would take one of three values corresponding to all the combinations of possible sensitive answers, i.e. 'YES' only, 'NO' only, 'YES' and 'NO'. Finally each of the $S_{3}(i)$, $i = 1, 2, \ldots k$, would take one of two values 1 and 2, which would indicate a 'positive' or a 'negative' sensitivity correspondingly.
2) 'Dressing the questionnaire up'. First the questions would be 'dressed up' with normal (= not sensitive) E.P.s. A random number generator (function) could be used to generate different numbers \( N_\tau \), all between 1 and \( N \), the total number of which corresponds to the 'optimum' amount of this variable ('sparse' variable as described in section 4.3).

Then, the random number generator would generate a number say \( N_e \) between \( F(C(2)+1) \) and \( F(C(3)) \), i.e. a number corresponding to one of the total of the E.P.s, and not only the different ones. The corresponding E.P. could then be specified in the file of phrases by its order in the \( F \) array say \( O \), which should obey to the relation \( F(O-1) < N_e < F(O) \). Then the specific E.P. could be located by using equations (1) and (2) for \( K=0 \), and the length of the particular E.P. \( L_3(O) \).

The question \( A \$ (N_\tau) \) and the corresponding length \( L_1(N_\tau) \) would then be updated as follows: The E.P. would be merged with the \( A \$ (N_\tau) \), and this would comprise the new \( A \$ (N_\tau) \), while the sum \( L_3(O) + L_1(N_\tau) \) would comprise the new \( L_1(N_\tau) \). This process would continue for all \( N_\tau \) 's.

Then, the questions would be 'dressed up' with C.P.s. This could be done in a similar way to that of E.P.s. Again a random number say \( N_\tau \), between 1 and \( N \), would be generated, the total of \( N_\tau \) 's corresponding to the optimum amount of this variable ('sparse' variable as described in section 4.3).

Similarly a second random number say \( N_e \) would be generated, corresponding to one of the C.P.s. This number should be between \( F(C(3)+1) \) and \( F(C(5)+2) \). Again the order of this C.P. in the array \( F \), say \( O \), would be needed, and this order would be defined by the relation \( F(O-1) < N_e < F(O) \). In this case, there are four possibilities for \( O \): a) \( O = C(4) \), i.e. the C.P. to be a pre-phrase; b) \( C(4) < O < C(5) \), i.e. the C.P. to be a post-phrase; c) \( O = C(5)+1 \), i.e. the case of an Q.I.P.; d) \( O = C(5)+2 \), i.e. the case of a I.T.Q.
The question $A_{\mathcal{Q}}(N_{T})$ and the corresponding length $L_{1}(N_{T})$ could then be updated in each of the above four cases as follows:

a) The pre-phrase would be merged with the question $A_{\mathcal{Q}}(N_{T})$, the former being specified by equations 1 and 2 (for $K=0$), and this would be the new $A_{\mathcal{Q}}(N_{T})$. The new $L_{1}(N_{T})$ would be the sum of $L_{3}(0) + L_{1}(N_{T})$.

b) The same process with a) would be followed, except that the question would now be merged with the post-phrases (i.e. the post-phrase would follow the question).

c) The question $A_{\mathcal{Q}}(N_{T})$ would be updated by following rules based on the recommendations of section 2.4.2 A) d), to transform a S.S.Q. to a I.T.Q. and vice-versa. The $L_{1}(N_{T})$ should also be updated.

d) Same process with c) would be followed, except that in this case the rules for the transformation of the S.S.Q. for Q.I.P. would be based on the recommendations of section 2.4.2 B).

The above process would be followed for all $N_{T}$'s, as for the E.P.s.

Then the 'sensitive' questions would be 'dressed up'. One way of dressing the 'sensitive' questions up would be to do it at this stage as follows. For each 'sensitive' question $A_{\mathcal{Q}}(S_{1}(i))$, a new version of the next question corresponding to the 'sensitive' answer (s) would be constructed. This new version of the question would be consisted of a 'sensitive' E.P. merged with the question following the $A_{\mathcal{Q}}(S_{1}(i))$, and depending on the values of $S_{2}(i)$ and $S_{3}(i)$. This new version would then be stored in an alphanumerical array $A_{1\mathcal{Q}}(i)$, $i = 1, 2, \ldots k$. Moreover, the 'sensitive' E.P. would be either 'positive' or 'negative' depending on the value of $S_{3}(i)$.
Let us take as an example the case for $S_2(i) = 1$ and $S_3(i) = 1$, i.e. the case of a 'YES' sensitive answer with a 'positive' E.P. required. In that case first a Positive Phrase would be picked up randomly from the file of phrases. For that a random number, say $N_e$, would be generated between 1 and $F(C(i))$, i.e. a number corresponding to a phrase among the total of all positive reinforcements (not only the different ones). The order of this phrase in the $F$ array, say $0$, would then be specified by the relation $F(0-1) \leq N_e \leq F(0)$ etc. along the lines of normal E.P.s. Then the positive reinforcement would be retrieved for the array of phrases by using equations (1) and (2) for $k=0$. This positive reinforcement would then be merged with the question $A \$ (N_1(S_1(i))) to become the new question $A_1 \$ (i)$. A similar process would be followed for the other possible values of $S_2(i)$ and $S_3(i)$.

For the 'new' questions $A_1 \$ (i)$, $i = 1, 2, \ldots k$, four vector arrays would be created, array $L_2(i)$, containing their length, and arrays $N_4(i)$, $N_5(i)$, $N_6(i)$, containing their logical connections with the rest of the questions. In this case for say a 'YES' only sensitive response, $N_4(i)$, $N_5(i)$, and $N_6(i)$ would be equal to $N_1(N_1(S_1(i)))$, $N_2(N_1(S_1(i)))$ and $N_3(N_1(S_1(i)))$ correspondingly, i.e. equal to the logical connections of the question which would appear after the $S_1(i)$ question in the original questionnaire, should a 'YES' response was given (to the $S_1(i)$ question).

It would be helpful to put all the questions together in an array $A_2 \$ (i)$, their length in an array $L_4(i)$, and their logical connections in three vector arrays $N_7(i)$, $N_8(i)$, $N_9(i)$, $i = 1, 2, \ldots N, N+1 \ldots N+k$. In this case, $A_2 \$ (i) = A \$ (i)$, $L_4(i) = L_1(i)$, $N_7(i) = N_1(i)$, $N_8(i) = N_2(i)$, $N_9(i) = N_3(i)$, for $i = 1, 2, \ldots N$, and $A_2 \$ (i) = $A_1 \$ (i)$, $L_4(i) = L_2(i)$, $N_7(i) = N_4(i)$, $N_8(i) = N_5(i)$, $N_9(i) = N_6(i)$, for $i = N+1, N+2, \ldots N+k$. The logical connections of the sensitive
questions A2$ (S1 (i)), i = 1, 2, ..., k, should also be updated as follows: if $S2(i) = 1$, i.e. a sensitive reinforcement would be required for a 'YES' response only, then $N7 (S1 (i)) = N+i$, if $S2 (i) = 2$ then $N8 (S1 (i)) = N+i$, and if $S2 (i) = 3$ then both $N7 (S1 (i)) = N+i$ and $N8 (S1 (i)) = N+i$.

Finally, the $N+k$ questions can be stored in a file, since to have them in an array would not be very economical in space, due to the fact that the array elements are usually of fixed length.

Assuming that this file would have records of fixed length (the usual case) of say Z characters each, then this file would contain $T$ records, where $T$ would obey the relation:

$$T = \sum_{i=1}^{N+k} \frac{L4(i)}{Z+H}, \text{ with } H = \begin{cases} 0 & \text{if } \sum_{i=1}^{N+k} \frac{L4(i)}{Z} = 0 \\ 1 & \text{otherwise} \end{cases}$$

Consequently the $L$th question would start from a record and a character within this record satisfying equations (3) and (4), which are similar to equations (1) and (2).

(3) Number of Record = $\sum_{i=1}^{L-1} \frac{L4(i)}{Z} + 1$

(4) Order of Character = $R \left( \sum_{i=1}^{L-1} \frac{L4(i)}{Z} \right) + 1$
Therefore, what we finally need to store from this questionnaire is the file with the questions, and the four vector arrays L4, N7, N8 and N9, all of N+k elements, while all the other intermediate arrays do not need to be stored.

These data structures are shown in figure 37.

12.4 Rules for the construction of the second part (stage 2) of the Computer System

By typing the particular command and the name of the questionnaire, the system would work in stage 2, i.e. carry out the interview.

After starting the system, it would start interacting with the interviewee. First, some introductory questions could be displayed on the V.D.U. along the lines of section 2.2.

Then the main questions would be presented. For an answer to question i, $1 \leq i \leq N+k$, next, question N7(i) would appear for a 'YES' answer, N8(i) for a 'NO' answer and N9(i) for a '?' answer. This process, starting from the very first question, would continue till one of these logical connections points to the end of the questionnaire. During the above described process, each question i would be located in the file of the particular questionnaire. At the end of each questionnaire, a thanks giving to the interviewee would be added, along the lines of the four programs which have been analysed.

For each interviewee and each particular interview questionnaire four arrays would be needed. One, say R (N+k), to store the questions asked, the second, say P (N+k), to store the answers to those questions, the third T1 (N+k), to store the reaction time for each question, and the fourth T2 (N+k) to store the total interaction time. The above four arrays, as shown in figure 38, would be printed at any time on request.
13. GENERAL DISCUSSION

This chapter contains a general discussion of the studies described so far. It refers to the main points of the analysis and of the experiments on interview acceptability and on accuracy of the derived information. (More detailed discussion of these items has been presented in sections 2.7, 9 and 11 correspondingly.) During this chapter the new parts of the research are highlighted, and the methods used and the conclusions drawn are discussed, particularly in relation to the work of other researchers in the field of man-computer interviewing. Some suggestions are then noted for developing further the computer software system; and recommendations are made for the maintenance of the system and for its further perfection, in order to become more adaptive to the individual user and to enable the exploitation of its full potential. Finally, the original hypotheses and aims, as described in section 1.3, are re-examined under the light of the research done to consider if and to what degree they have been met.

The studies and experiments of this research programme all have the general purpose of understanding more fully the computer interview process. As a consequence it may well be possible to automate further the computer interview, while at the same time satisfying the need of people for 'natural' and 'friendly' interaction.
Therefore, this research had three specific aims:
1. To investigate the linguistic context of an acceptable computer interview questionnaire.
2. To construct a computer system, by using the findings of the linguistic studies and experiments, which would have the properties described in section 1.3 and would carry out interviews acceptable to people.
3. To achieve the second aim without entering the full complexity of the semantics of the language.

13.1 Summary of the Analysis

The first part of the studies consists of a grammatical, syntactic and semantic analysis of four history-taking interviewing questionnaire programs, which were written by Dr. Evans and his team at N.P.L. London. The reason for this analysis was to study the linguistic context of computer interviewing questionnaire programs already shown to be highly acceptable to the interviewees.

The four medical history-taking interviewing programs were chosen, because the 'friendly' and 'chatty' way in which they put the questions was a major factor for their high acceptability to actual patients (as described in section 1.1.2).

This analysis of the linguistic context of an acceptable computer interview questionnaire program is research which is believed to have been carried out for the first time in this study. The method used during the analysis and the main findings are as follows.

First, the Simple and Straightforward Questions (S.S.Q.s), that is the shortest and simplest questions having the same meaning, were extracted from the questions of the four computer questionnaires.
Then, the several phrases added, and the alternative ways of asking the questions, both aiming to increase the variability of the interaction and make it more 'natural' and 'friendly', were identified. Moreover, simple transformational rules, based on the elements of the surface structure of the questions, were stated, enabling the transformation of most of the questions differing from S.S.Q.s to S.S.Q.s and vice-versa (section 2.4).

These 'friendly' and 'chatty' phrases added to the S.S.Q.s and the alternative ways of asking the questions, were called Linguistic Style Elements (L.S.E.).

Then, the L.S.E. were classified into several categories: 1) phrases preceding the question; 2) the question itself, if different from S.S.Q.s; 3) phrases following the question, which were further classified into a number of sub-categories.

Among the L.S.E., about 70% were found to be context-free and thus directly usable by the proposed system. Then, the pattern, which the L.S.E. follow among the questions, was examined. By using statistical tests, it was found that the L.S.E. do not seem to follow any common pattern among the questions, and that randomness is the most probable case.

Finally, to test the 70% seemingly context-free L.S.E. further, if they bear any relation to the meaning of the questions, four variables, very important to the interview situation, were defined. Yet, no correlation between the L.S.E. and these four variables was found.

13.2 The Experimental Methods

The second part of this research consists of six experiments, aiming to test the usefulness of the findings of the previous analysis in a computer interview of a general subject, since the L.S.E. would be expected to be valid for an interview of any kind.
This second part can be split further into two sub-parts, one including measurements of people's attitude, the second consisting of measurements of reproducibility and accuracy, both in a computer interview and in a human interview.

The attitude of people towards the computer and human interviews was assessed by interviewing people by computer (three experiments described in chapters 4, 5, 7), and by human interviewers (one experiment described in chapter 6), and by both computer and human interviewers (one experiment described in chapter 8).

During the computer interviews, the computer was presenting the questions in natural English on a V.D.U., while the interviewees were answering the questions at their own pace by pushing one of three buttons on a keyboard marked 'YES', 'NO' and '?'.

During the human interviews, the experienced interviewers were either instructed to follow the exact wording of the written questions (one experiment described in chapter 6) or they were allowed to introduce their own linguistic style when asking the questions (two experiments described in chapters 8 and 10).

The method used after all interviews, computer or human, to measure people's attitude towards the interview was devised by interviewing by computer 54 people, and it was based on objective psychological scaling techniques and mathematical models, i.e. the Semantic Differential technique and the Factor Analysis method. The final device was comprised of six bipolar linearly related adjectives, which constitute a factor called 'Personal Evaluation'.

The reproducibility and the accuracy of the information derived during the computer and human interviews were assessed by interviewing people by both the computer and human interviewers.
Then, the number of disagreements with the consensus was used as a measure of the precision of the information derived by each interviewer, while error rates were assessed by a statistical model by Dawid and Skene, based on the maximum likelihood technique and a computational algorithm.

Comments on the Method used to Measure People's Attitude towards the Interview

Among the published studies, aiming to assess attitudes towards the computer interview, most involve direct questioning or short questionnaires, both producing biased scores, while only a small proportion of studies involve the use of recognised psychological scaling techniques (e.g. Franklin et al., 1967; Hulka et al., 1970).

The latter consist mainly of 'Thurstone' type scales, or the Semantic Differential (S.D.). The first group of methods are rather tedious, and costly in time and effort. On the other hand the S.D. is quicker and easier to construct, at an equivalent degree of validity (Lucas, 1976).

Yet, to the best of our knowledge, no study appears where the S.D. has been combined with the Factor Analysis method to result in a practical, very quick and easy to use, still reliable device to measure attitude towards the computer interview (i.e. only six bipolar adjectives related through a linear relation, the latter being decided by applying mathematical models to the experimental data).

Comments on the Methods used to Measure Reproducibility and Accuracy of the Information

The number of disagreements with the consensus is a 'good' indicator of observer's precision, stable enough even to removal of individual cases, and it has been used in several studies for this purpose (e.g. Lucas et al., 1977).
The statistical model used to estimate accuracy is very recent and makes possible a method of analysis which previously suffered from severe computational constraints.

The validity of the above two methods in relation to the particular design used in the relevant experiment is also discussed in detail in chapter 11.

13.3 The Results

The results of the experiments on people's attitude and accuracy and precision of the derived information can be split into two groups. The first group refers to comparative measures of the acceptability and accuracy of several versions of computer interview questionnaires, differing in the amount and pattern of the L.S.E. used. Moreover, one experiment was conducted to study the effect of L.S.E. on the acceptability of human interviews to the interviewees. The second group of results refers to how the 'best' computer interview questionnaire, i.e. the one incorporating the optimum amount and pattern of L.S.E., is compared with human interviews, as far as their acceptability to the interviewees, and the precision and accuracy of the derived information are concerned.

Results on the Effect of L.S.E. in Computer and Human Interviews

The main results of several tests on the usefulness of L.S.E. (first group of results) are as follows.

First, random Encouragement and Chattiness (these two variables representing the L.S.E.) seem to have a significant effect on interviewee's acceptance of computer interviews, but not on the reproducibility and accuracy of the derived information.

Second, there appears to be an optimum in the amount of random Encouragement and Chattiness used in a computer interview.
Third, the random pattern of L.S.E. of optimum density seems to be as acceptable to the interviewees as more planned patterns, and as those patterns devised by professional interviewers aiming to maximize the acceptability of the particular computer interview to the interviewees.

Fourth, while random Encouragement and Chattiness seem to have a significant effect on interviewee's acceptance of computer interviews, this does not appear to be the case for human interviews. In particular the variable Chattiness seems to have no effect on interviewee's acceptance while the variable Encouragement might even have a 'negative' effect, with more random Encouragement resulting in less acceptable versions.

The fact that random patterns of Encouraging and Chatty phrases proved acceptable in the present studies cannot necessarily be taken to imply that this would be the case for all types or lengths of computer interview. Therefore these results are valid for short interviews, i.e. about 10-15 minutes, and questionnaires of a non-sensitive subject only. A possible interpretation of these results would be that the subjects, during the short computer interviews (about 10 minutes), experienced a dialogue which effectively simulated a man-man interaction situation; in this they would probably accept random Encouragement and Chattiness during its first exploratory stages, but would possibly require a more planned pattern of those variables as the interaction goes on and the level of knowledge and understanding of the interacting members grows. Furthermore, the optimum amount of Encouraging and Chatty phrases during the computer interviews was derived from and resembles a man-man interaction situation, where a person uses that format probably by intuition.

Another aspect of this question is perhaps indicated by the 'neutral' or 'negative' effect of random Chattiness and Encouragement found in the attitude expressed by interviewees during human interviews. These results in human interviews and the previous results in computer interviews possibly indicate that the interviewees can accept a computer (a machine) interacting like an imperfect human, but cannot accept this from a human.
The results summarized above have also answered an important question, i.e. whether differences exist in the reproducibility and accuracy of the information derived by computer interview questionnaire versions, differing in their acceptability to the interviewees or not. Yet, it is important to notice that the negative answer given to this question by the present studies, i.e. that there is no difference in the reproducibility and accuracy of the above versions, applies on the questionnaire about leisure time and hobbies, which was used during all six experiments conducted, and perhaps applies on every questionnaire of a general and non-sensitive subject. However, the case might be different if the people were interviewed on a sensitive subject, e.g. personal relationships. For the latter case it is possible that the more acceptable version would elicit more accurate information too. Therefore the author would welcome any research aiming to clear this interesting point.

Results on the Validity of the Computer Interview - Comparison with Human Interviews

The results of several tests on the validity of the computer system, incorporating the information gained from the previous analysis, (second group of results), can be summarized as follows. Computer interviews may be as acceptable to the interviewees, and as accurate as human interviews, while they are at least as precise (reproducible) as the latter. The important implication of the above results is that in principle the computer could be used instead of the human specialist to collect information without loss of accuracy. This would not only save human time and effort by avoiding the routine job of interviewing people, but would also provide all the information stored directly in the computer and ready for further handling.

However, although the combined scores of acceptability of the above results did not differ for the two methods (computer and human interviewing), one experiment has indicated that computer interviews may be significantly more interesting, but less friendly, than human interviews.
These two indications do not appear to have been referred to in any study before. The former can be interpreted as the influence of the novelty of the situation (computer interview) to most subjects, and as a possible indication of the users' acceptance of the computer interview as a useful and time saving application of computers in real life problems.

The latter probably reflects the fact that the adjective 'friendly' is a word closely related to human beings, and therefore it is not very well comprehended when used for a computer, an inanimate machine.

Finally, on the question of possible sex differences in the attitude scores no significant differences of this kind were generally obtained, either for the computer or for the human interviews. These results possibly indicate that the two sexes have no different attitudes towards computer and mainly the small sample size was not sensitive enough to measure such differences should any exist.

13.4 The Computer Software System

Reviewing these research studies in the context of the general framework of computer interviews, the following comments can be made.

A computer system has been built, based on the findings of an analysis on the linguistic context of four computer interviewing questionnaire programs, to carry out 'natural' and 'friendly' interviews.

This system is able to carry out an interview of any kind, should it be fed with the relevant questionnaire and the subsequent branching instructions, while at the same time satisfying the need of people for 'natural' and 'friendly' interaction.

These constitute a further step towards the automation of the interview process, through the replacement of the programmer, who had up to now to program every single questionnaire program for each particular subject separately.
Limitations

One limitation of the system, imposed on it by its own design, is the requirement for the input to be in the form of the basic questionnaire of Simple and Straightforward Questions, together with the relevant branching instructions for every single subject about which an interview would be required.

The former is not a serious limitation, since for most of the frequently conducted interviews questionnaires with checklists already exist. For example, for medical history-taking interviews, questionnaires of S.S.Q.s for almost every branch of medical science are available in the hospitals. The latter (logical connections among the questions or branching instructions i.e. the number of the next question after a possible 'YES', 'NO' or '?' answer) is very simple, and a specialist would not be required to specify them.

Another limitation of the current implementation of the system is the limitation on the way the interviewee can communicate with the system, e.g. small number of possible responses, no way to correct a possibly faulty response, etc. This profile of the system was deliberately restricted, in order to test the usefulness of the findings of the analysis on which the system was based under conditions similar to those of the four computer interviews which were analysed. Yet, these restrictions would not be difficult to overcome by altering the programming implementation of the system, should the need arise.

Finally, another limitation of the complete automation of the system, as already mentioned in this study, is that positive reinforcements, like congratulating phrases, could only follow an answer of the interviewee showing a personal success or a social advantage. Similarly phrases like 'I am sorry to hear that' ('negative' reinforcements) would suit only answers of the interviewees showing a personal failure, or a social disadvantage.
Therefore, because of the above restriction imposed by a small percentage of phrases (about 3% of the total of the Encouraging Phrases used), it would be necessary for a human to indicate the relevant 'sensitive' questions-answers during the stage of feeding the system with each particular questionnaire of S.S.Q.s. This is a straightforward job, and again it would not require a specialist to do it.

Then the process would be completely automatic, and the computer system would be able: a) to 'dress up the S.S.Q.s' to become more 'natural' and 'friendly' (using the information gained from the analysis), and b) to carry out the interview at a later stage using the revised questionnaire. Moreover during the interview the system would also be able to keep trace of the number of the questions asked, the answers given, the interviewee's reaction times, etc. (the reaction time would be very important during special types of interviews, e.g. it would provide valuable information about sensitive questions in psychiatric interviews and this, i.e. the capability of computers to measure reaction times, is not a limitation of the system, but an important advantage of it and of computer interviews in general).

Evaluation - Possibilities for the Future

The computer interviewing system was evaluated for acceptability to the interviewee, and accuracy and reproducibility of the derived information.

The positive results of the tests on the system's acceptability to the interviewees strengthen further the similar findings of other studies, i.e. that computer interviews, if properly designed, may be as acceptable to the interviewees as human interviews. This has been indicated by several studies, either on medical history-taking interviews (e.g. Mayne et al., 1969; Evans et al., 1973; Lucas et al., 1977b) or on computer interviews in general (e.g. Hedl et al., 1971; Johnson et al., 1971).
On the reproducibility and accuracy of the information derived during the computer interviews, this system has also been successful, by being as reproducible and as accurate as human interviewers (the validity of the methods used is discussed in detail in chapter 11).

These results also strengthen further similar findings of the studies on medical interviewing systems (e.g. Evans et al., 1971; Card et al., 1974; Lucas et al., 1977a), on interviews in computer administered tests (e.g. Hitti et al., 1971; Lushene et al., 1974), and on computer assisted counselling (e.g. Melhus, 1971; Price, 1974).

On the monetary cost, the suggestions of Lucas et al. were taken into account (Lucas et al., 1977a), about the use of a mini-computer to conduct the computer interview, which they found to be cheaper than the specialist interviewer, while the new technology of micro-processors is a cheaper possibility for the future.

This computer system was mainly designed to test the effect of the L.S.E. on interviewee's acceptability. Therefore it was not made to be adaptive to the individual user, since this would not be necessary for the above tests. On the question of adaptation of the system the suggestions by Lucas (Lucas, 1974), for the adaptation of the speed of presentation of the questions, and of the number of responses available to the user, depending on the age and the intelligence of the interviewees, seem relevant. Moreover the E.P.s and C.P.'s could be used adaptively too, depending on individual characteristics like age, intelligence etc. Yet, extensive experimentation would be required to establish the correlation between the patterns and amount of E.P.s and C.P.'s, and the above mentioned characteristics in a successful computer interview. It is also possible that this correlation would be different for interviews of a 'neutral' and a 'sensitive' subject, or for interviews of different length, and it would be interesting to study 'sensitive' and 'longer' interviews to see if they are more subtle and sophisticated, and also to study the relation of L.S.E. with the above mentioned individual characteristics.
Finally, it is the strong belief of the author that the full potential of this system cannot be realised, unless this system is incorporated in a larger system, which would be able to handle the elicited information by statistical decision theory models, to calculate probabilities and decisions. For example, for medical interviews, the system could be part of a larger system, which would be able not only to take the medical history of patients, but also to calculate diagnostic probabilities for certain diseases, along the lines suggested by Card and Knill-Jones (Card, 1967; Knill-Jones et al., 1973).

13.5 Suggestions for the Maintenance and further Perfection of the Computer System

Following from the discussion in the previous section some further comments and suggestions seem appropriate. First, it is noticeable that the number of available answers for the user could easily be increased, without much programming effort and space requirements, providing that it does not exceed a certain limit, say the number 10. On this Lucas (1976) suggests the addition of the certainly YES/NO, possibly YES/NO and probably YES/NO answers to the YES, No, '?' ones. Moreover, the same author recommends an adaptive use of those options, depending on the group of age and intelligence of each particular individual. For the latter (I.Q.) he suggests that it could be estimated from the reaction times of each subject to the preliminary questions (he has found a negative correlation between I.Q. and reaction times). These suggestions seem attractive, though more experiments to test this correlation would be essential.

Another suggestion, on the adaptation of the system to the individual, would be to adapt the amount of Chattiness and Encouragement, depending on the user's individual characteristics (e.g. age, intelligence, etc.). This is a possibility which would require extensive experimentation to devise the patterns of the correlation of these variables in an acceptable computer interview.

The above suggestions could be called 'short term' adaptive component of the system.
Then the 'long term' adaptive component would aid the updating of the questions, and of the logical connections of the questions of each stored questionnaire. Towards this direction, the user's reaction times, the number of '?' responses, and the proportion of times each question was asked could be used as indicators of the usefulness of the questions. Then, depending on these indicators, which could be printed on request, each 'troublesome' question could be updated or completely eliminated. Yet, this updating or elimination of questions could not be done without entering into the semantics of the language, i.e. without human involvement.

Further, it would be very useful if all this information about the questions of all the questionnaires stored in the system were stored in a data base, to which the system would be linked.

Even more important would be to link the system with a data base to store the information about each particular interviewee (providing that the issue of the privacy of the stored information would be settled in such a way as to satisfy the interviewees). For example, for medical interviews it would be essential to keep the medical records of all patients. Therefore, it would be very helpful to link such an interviewing system to a data base, which would store and update the information about each patient after a computer medical interview.

But the most important of all, especially for the appreciation of the full potential of such a system in the near future, would be to enlarge it in order to contain statistical decision theory models, to handle the elicited information by calculating probabilities and decisions. For example, for the medical interviews, this would result in calculations of diagnostic probabilities for a set of illnesses, possibly along the lines suggested by Card (1967) and Knill-Jones et al. (1973).
13.6 Limitations of the Research and of the Results

There are two major restrictions of this research which impose essential limits to generalisation of the results. One is the short length of interviews (about 10 to 15 minutes), the other is the non-sensitive subject of it. For example it is very likely that the results would be quite different for a longer interview, where the structure of E.P.s and C.P.s would be expected to be more subtle and sophisticated, and similarly for an interview of a 'sensitive' subject, for which the E.P.s and C.P.s could possibly present much more complicated patterns. Therefore it seems fair to say that the results of this study are valid for a short interview (about 10-15 minutes) and for a non-sensitive subject only, while it would be unsafe to generalize beyond those limits.

Within the above limits there are a few more questions which could impose further restrictions on the results of this research. One such question is the validity of the device used to measure people's attitude towards the interview. For example someone could question the accuracy of the final device (consisting of 6 bipolar adjectives) to measure attitudes in experiment 1 and the other five experiments, since it was devised from an experiment where the 6 adjectives were presented with 14 more adjectives, but it was used in the other 5 experiments where the six bipolar adjectives were presented without any others. Finally, the methodological issue of the within-subjects design used during two experiments, one to compare computer and human interviews, the other to measure reproducibility and accuracy of information during interviews, could raise some objections too.

The first question on the validity of the final device to measure attitudes was answered satisfactorily by comparing the attitude scores of experiment 1 with those of the other five experiments (for the same questionnaire version). The lack of significant differences between these two groups of scores suggested that there was no significant interaction between the 6 adjectives and the remaining 14 ones in experiment 1, and therefore no significant difference of the accuracy of the final device during the 6 experiments.

The within-subjects design of the experiment
to compare the acceptability of computer and human interviews was chosen because it is more sensitive, although less reliable than the between-subjects one, and therefore would detect more easily significant differences between the two methods (i.e. using less subjects). These significant differences, if any were found, could then be validated further by using a between-subjects design. However since the within-subjects experiment showed no significant differences between the two methods, the between-subjects one was no longer necessary, without this to impose any limitation on the results.

As far as the within-subjects design to measure reproducibility and accuracy of information this is a restriction of the existing statistical models, since there is no statistical model, to measure accuracy of information, based on a design other than the within-subjects one. Moreover some points related to this model, and especially to the time intervals between the interviews of the same subject, have been extensively debated in chapter 11. In conclusion it seems that within the limits of the model meaningful results were obtained both of the relative accuracy and reproducibility of several versions of different acceptability to the interviewees, and of the relative reproducibility and accuracy of the two methods, i.e. human and computer administered interviews.

Summarising it could be said that the results presented so far and the conclusions drawn could be considered as valid within the limits of the short length of the interview (about 10 minutes) and for 'neutral' questionnaire only, since only short interviews and a 'neutral' questionnaire were used during this research.

13.7 A Revised Formulation of Computer Interviewing

Three years ago, when this research started, computer interviewing did not give much emphasis on the style of the questions, or the way the questions were phrased. Moreover, the few pioneers who had realised the importance of the
style of the language used in a computer interview (e.g. Dr. Evans from the National Physical Laboratory) had to adapt their patterns of questionning, separately for each particular questionnaire subject, by intuition and personal style.

This research has not only studied in depth the linguistic context of an acceptable computer interviewing questionnaire, but also has proposed and has developed the first exploration of a new formulation for computer interviewing. The essence of this new formulation is to find out by evaluative research the patterns of the linguistic variables which constitute an acceptable computer interview, and then to use the computer itself to reconstruct these patterns and carry out the interview. Therefore this new formulation requires both extensive research on intuitive patterns of phrases used in acceptable interviews and the setting up of 'computer compatible' rules, i.e. rules which can be implemented in a software system; thus the computer is enabled to convert a questionnaire of simple and straightforward questions into an interview program and then to carry out interviews in a 'natural and friendly manner'. As a result the only requirement of this formulation would be to feed the computer system with questionnaires of straightforward questions and the logical structure connecting them (i.e. the number of the next question after a certain answer to a question among a discrete number of choices).

The benefit of this new formulation is that it greatly reduces the tedious and routine work of the programmer and the specialist e.g. doctor, psychologist etc., who had up to now not merely to combine their skills on the essential questionnaire content and logic, but also to do much detailed writing to produce an acceptable computer interview for each particular questionnaire subject. Therefore this formulation offers a further step towards the automation of computer interviewing, while still satisfying the need of people for natural and friendly interaction.

This formulation has been applied so far, in this
research programme, for short interviews (i.e. 10 to 15 minutes long) and for 'neutral' questionnaires only. The relevant rules which have been developed are described in Chapter 12. Yet it seems reasonable to suggest that the formulation could be generalised for longer interviews and 'sensitive' questionnaires. To do this, further studies would be needed on the relevant patterns of E.P.s and C.P.s, and on the computer compatible rules needed for this case also, as for the case of short interviews on a 'neutral' subject. Further, the pool of phrases (E:P:s and C.P:s) used during this study (shown in Appendix 3) could be extended by studying other acceptable computer interviewing questionnaires.

However this revised formulation cannot be safely accepted as 'the' new formulation of computer interviewing until successfully tested on other kinds of interviews and in other subject areas. It is hoped that this research, and the suggestions for extending it given above, will lead to further constructive developments in the field of man-computer interviewing.

13.8 Have the Original Aims been fulfilled and to what Degree?

Examining the original aims of these studies, it can be said that the first aim has been successfully completed through the analysis of four medical history-taking interviewing programs, described in chapter 2. As far as the second aim is concerned, i.e. the construction of an automatic computer system to conduct 'acceptable' interviews for any subject, it can be said that this aim has been completed to a high degree too (relevant experiments are described in chapter 4-11).

Yet, the above described studies have indicated that the computer system cannot be completely automatic, as it was originally ambitiously desired. Therefore, the presence of a human, to 'look at' the questions before feeding the system, was found necessary, even though this would be a
straightforward and not at all time-consuming job, for which a specialist would not be required. Thereafter the process would be completely automatic, and the system would be able to conduct 'natural', 'friendly' and 'accurate' interviews, as originally aimed.

The third aim of this research, i.e. to construct the acceptable computer interviewing system without entering the full complexity of the semantics of the language, has been fulfilled too. This is because a) the hypothesis that the context-free L.S.E. may have a significant effect on interviewees' acceptance of computer interviews has been experimentally supported, and b) experimental results have also shown that a certain pattern and density of context-free L.S.E. results in a computer interview questionnaire version, which is as acceptable to the interviewees and as accurate as human interviews.

The fact that the context-free L.S.E. may have a significant effect on people's attitude towards computer interviews support also indirectly the original hypothesis underlying this research, i.e. that the L.S.E. (linguistic factor) has a significant effect on people's acceptance of computer interviews.

13.9 Conclusions

In conclusion it can be said that the computer system constructed, although slightly restricted from its original ambitious aims, can be seen as a considerable improvement towards the automation of the computer interview process, while still satisfying the interviewees' need for 'natural' and 'friendly' interaction.

Moreover, the analysis on the linguistic context of four highly acceptable interviewing programs, and the studies of the relationship of the derived linguistic style elements to the acceptability of interviews, together comprise a new contribution to science in the field of man-computer interviewing.
14. CONCLUSIONS

From the studies and experiments on computer interviewing described so far, the following may be suggested in conclusion, but it must be remembered that the results may not yet safely be generalised beyond the duration and length of interview studied (about 10 minutes and 25 questions) or beyond the neutral non-emotional contents of the interviews used in this research.

1. An analysis of 4 highly acceptable medical history-taking interviewing programs has shown that they consist of S.S.Q.s and phrases differing from S.S.Q.s, which were called L.S.E.

2. The L.S.E. were then classified into several categories and sub-categories.

3. Among the L.S.E. about 70% were found to be context-free, and thus directly usable by a computer system to carry out interviews acceptable to the users.

4. Several tests have failed to find any pattern that the L.S.E. follow, and it seems that they follow a random pattern among the questions.

5. During an experiment on computer interviewing a method to measure people's attitude towards the interview has been devised (developed experimentally) in the form of six linearly related bipolar adjectives using the semantic differential scale technique.

6. Random Encouragement and Chattiness (these two variables representing the L.S.E.) had a significant effect on interviewee's acceptance of computer interviews.

7. There appears to be an optimum in the amount of random Encouragement and Chattiness used in a computer interview.

8. During a computer interview, the random pattern of L.S.E. of optimum density may be as acceptable to the interviewees as more planned patterns. Moreover, it may be as acceptable as those patterns which were devised by professional interviewers aiming to maximise the acceptability of the particular computer interview to the interviewees.
9. Random Chattiness had no effect on interviewee's acceptance of human interviews, while random Encouragement might even have a negative effect, with more random Encouragement resulting in less acceptable versions.

10. During computer interviews, there was no difference in the reproducibility and accuracy of the information derived by several questionnaire versions, differing in the amount of random Encouragement and Chattiness (and in their acceptability to the interviewees).

11. Computer interviews may be as acceptable to the interviewees, and as accurate as human interviews, while they are at least as reproducible as the latter.

12. Computer interviews may be more interesting, but less friendly than human interviews.

13. Rules have been stated for the construction of a computer system to carry out 'natural' and 'friendly' interviews of any subject. These rules were independent of a particular computer language. The resulting system could be seen as an improvement towards the automation of the interview process, through the replacement of the programmer, who has up to now had to write every single questionnaire program for each particular subject separately.

14. It is the strong belief of the author that the full potential of such a system cannot be realised, unless this system is incorporated in a larger system, which would be able to handle the elicited information by statistical decision theory models, to calculate probabilities and decisions. For medical interviews this system would be able not only to take the medical history of patients, but also to calculate diagnostic probabilities of certain groups of diseases and present both sets of data to the physician as aids to diagnosis and decision.
REFERENCES


Card, W.I., 1967, 'Towards a calculus of Medicine', Medical Annual, p.9, Bristol, John Wright and Sons Ltd.


Cliff, N., 1959, 'Adverbs as multipliers', Psychological Review, 66, pp.27-44.


Committee on Public Accounts, 1976, Report No. 6, H.M.S.O., para 118.


Fuller, D.J., 1977, 'An experiment to determine patients acceptability of an Automated Ante-natal history-taking Program', Report in print, Computer Science Division, National Physical Laboratory, Teddington, U.K.


Howe, E.S. 1966a, 'Verb tense, negatives and other determinants of the intensity of evaluative meaning', Journal of Verbal Learning and Verbal Behaviour, 5, pp.147-155.

Hulka, B.S., Cassel, J.C., Zyzanski, S.I. and Thompson, S.J., 1970, 'Scale for the measurement of attitudes towards physicians and primary medical care', Medical Care, 8, p.429.


Peter, E. and Donnan, H., 1972, 'A system for computer dissemination of occupational information', The School Counsellor, 19, pp.271-274.


Spiliotopoulos, V., 1975b, 'An Analysis of Four Medical Interviewing Programs', HUSAT Memo No. 106, Department of Human Sciences, University of Technology, Loughborough, U.K.

Spiliotopoulos, V., 1976, 'Studies and Experiments Towards a Computer Interview Acceptable to the User'. HUSAT Memo No. 130, Department of Human Sciences, University of Technology, Loughborough, U.K.


BIBLIOGRAPHY


APPENDIX 1 - FIGURES
The man hit the ball.

Figure 1.
**Program Pregnancy**

I = Introduction
E.P. = Encouraging Phrase
Pr.P3 = Freely attached or detached pre-phrase
Pr.P4 = One of the rest of pre-phrases
S.S.Q. = Simple-straight question
I.T.Q. = If-then question
Q.I.P. = Question with included (or embodied) phrase
D.Q. = Descriptive question
Po.P1 = Post-phrase explaining part of the question
Po.P2 = Post-phrase giving directions to the user

(P), (R), (Q), (W), (X), (S), (T), (Y), (V) = Repeated parts of the program

C = Continuation point
* = Connection point

![Diagram](Image)
PROGRAM PREGNANCY

Figure 2b
PROGRAM GASTRIC ULCER

The meaning of the used symbols is the same as in PREGNANCY.

Figure 3
PROGRAM SEX

The symbols as in Pregnancy
The symbols as in Prenumery. (X) A repeated part of the program.
Figure 14

Sex (Females)
Figure 17

Confidentiality

Question Number
Figure 19

GASTRIC ULCER

Arbitrary units

Question Number
Figure 20

GASTRIC ULCER

Total number of words of each question

Question number

10 20 30 40 50 60

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
Figure 25

Sensitivity vs. Total Number of Words of Each Question
Figure 27

GASTRIC ULCER. TOTAL NUMBER OF WORDS OF EACH QUESTION - SUBJECTIVITY
# LEISURE TIME AND HOBBIES

1. During week-days, Monday to Friday, is there any time which you have completely free of work to spend as you wish - that is leisure time? [Yes] [No] **Skip to Q3**

2. Do you have at least one full hour each week-day to spend as you wish? [Yes] [No]

3. Would you like to have more time than you have now on week-days for leisure activities? [Yes] [No] **Skip to Q6**

4. Would you like to have 2 hours or more, extra, each week-day for leisure activities? [Yes] [No]

5. Do you think that all working people in England should have more leisure time? [Yes] [No] **Skip to Q7**

6. Do you think that working hours in this country should be increased? [Yes] [No]

7. Does your job ever entail you working on Saturdays or Sundays? (for a student job has the meaning of studies) [Yes] [No] **Skip to Q10**

8. Do you work most weekends? [Yes] [No]

9. When you work on Saturdays or Sundays do you usually work full-time? [Yes] [No]

10. Do you have at least one full day at weekends to spend as you wish? [Yes] [No]

11. Do you have any hobbies, that is something you particularly like to do during your leisure time? **Skip to Q11**

12. Do you have more than one hobby? [Yes] [No]

13. Is your (main) hobby expensive to pursue? [Yes] [No]

14. On the average do you spend £5 or more each week on your hobby? [Yes] [No]

15. Do you consider that you spend a lot of your leisure time on your hobby? [Yes] [No] **Skip to Q17**

16. Do you usually spend 3 evenings or more each week on your hobby? [Yes] [No]

17. Does your hobby involve travelling? [Yes] [No] **Skip to Q19**

18. Do you usually use public transport for this purpose? [Yes] [No]

19. Do you do any voluntary work during your leisure time? [Yes] [No] **Skip to Q21**

20. Does this voluntary work involve you personally in any expenditure? [Yes] [No]

21. Do you take part in any sports during your leisure time? [Yes] [No] **Skip to Q23**

---

Figure 29a
22. Do you prefer quieter activities to taking part in sports?  
   Yes  No

23. Are there any hobbies, which you do not do now, but would like to do?  
   Yes  No  Skip to end

24. Is this because of the expense involved?  
   Yes  No

25. Is this because it would be very time consuming?  
   Yes  No

End
Each average is in the middle of a part of a line of one standard deviation length.
I think the interview was:

<table>
<thead>
<tr>
<th>GOOD</th>
<th>SLIGHTLY</th>
<th>SLIGHTLY</th>
<th>QUITE</th>
<th>EXTREMELY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAD</td>
<td>EXTREMELY</td>
<td>QUITE</td>
<td>SLIGHTLY</td>
<td>SLIGHTLY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERESTING</th>
<th>BORING</th>
<th>TENSE</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>UNFRIENDLY</th>
<th>FRIENDLY</th>
<th>UNPLEASANT</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>PLEASANT</th>
<th>UNPleASANT</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>KIND</th>
<th>CRUEL</th>
</tr>
</thead>
</table>

Figure 31
EACH AVERAGE IS IN THE MIDDLE OF A PART OF A LINE OF ONE STANDARD DEVIATION LENGTH.
Averages and Standard Deviations as in Figure 30.
Figure 34

Averages and Standard Deviations as in Figure 30
LEISURE TIME AND HOBBIES

Q1. During week-days, Monday to Friday, is there any time which you have completely free of work to spend as you wish - that is leisure time?

Q2. Do you have at least one full hour each week-day to spend as you wish?

Q3. Would you like to have more time than you have now on week-days for leisure activities?

Q4. Would you like to have 2 hours or more, extra, each week-day for leisure activities?

Q5. Do you think that all working people in England should have more leisure time?

Q6. Do you think that working hours in this country should be increased?

Q7. Does you job ever entail you working on Saturdays or Sundays? (For a student job has the meaning of studies).

Q8. Do you work most weekends?

Q9. When you work on Saturdays or Sundays do you usually work full-time?

Q10. Do you have at least one full day at weekends to spend as you wish?

Q11. Do you have any hobbies, that is something you particularly like to do during your leisure time?

Q12. Do you have more than one hobby?

Q13. Is your (main) hobby expensive to pursue?

Q14. On the average do you spend £5 or more each week on your hobby?

Q15. Do you consider that you spend a lot of your leisure time on your hobby?

Q16. Do you usually spend 3 evenings or more each week on your hobby?

Q17. Does your hobby involve travelling?

Q18. When travelling do you usually use public transport?

Q19. Do you do any voluntary work during your leisure time?

Q20. Do you give money to charities or other voluntary organisations?

Q21. Do you take part in any sports during your leisure time?

Q22. Do you prefer quieter activities to taking part in sports?

Q23. Are there any hobbies, which you do not do now, but would like to do?

Q24. Is this because of the expense involved?

Q25. Is this because it would be very time consuming?

End
DATA STRUCTURES FOR E.P.s AND C.P.s

File with E.P.s and C.P.s

<table>
<thead>
<tr>
<th>Record 1</th>
<th>P1$</th>
<th>P2$</th>
<th>L3(I)</th>
<th>C(5)</th>
<th>F(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record 2</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 36

I
I
I+1
m=I+2
### DATA STRUCTURES FOR QUESTIONS

**File with Questions**

<table>
<thead>
<tr>
<th>Record 1</th>
<th>Q1$</th>
<th>Q2$</th>
<th>L4(N+K)</th>
<th>N7(N+K)</th>
<th>N8(N+K)</th>
<th>N9(N+K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Record 2</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Record T</td>
<td></td>
<td>IQN+K$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 37
THE FOUR ARRAYS FOR EACH INTERVIEW

<table>
<thead>
<tr>
<th>R(N+K)</th>
<th>P$(N+K)</th>
<th>T(N+K)</th>
<th>T2(N+K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N+K</td>
<td>N+K</td>
<td>N+K</td>
<td>N+K</td>
</tr>
</tbody>
</table>

Figure 38.
APPENDIX 2 - THE COMPUTER SYSTEM
HELO38-2

CHARGE RATE LOW

PASSWORD:

RSTS V04B-17 LUT CS SYS#9 JOB 1 KB11 20-AUG-76 03:33 PM

DEPOSIT 333 CU

SPENT THIS WEEK: LOG 11 CON 909 CPU 8 DISK 76

NEW OR OLD-OLD MEDINA

READY

LIST

MEDINA 03:33 PM 20-AUG-76

7OPEN 'KFILE' AS FILE 9:DIM #9,K%(0),N8S(20)

9P15='MEDI1':P25='MEDI2':P35='CATALOGUE':P45='LIST':P65='DESTROY'

10PRINT:PRINT 'ANY REQUEST?'

:PRINT 'PLEASE FOLLOW THE ORIGINAL INSTRUCTIONS.'

:INPUT Z$

:IF INSTR(1,Z$,P15)>0 THEN 200

14IF INSTR(1,Z$,P25)>0 THEN 600.

15IF INSTR(1,Z$,P35)>0 THEN 30

16IF INSTR(1,Z$,P45)>0 THEN 60

18IF INSTR(1,Z$,P65)>0 THEN 180

19PRINT 'I AM SORRY BUT I CANNOT UNDERSTAND THIS REQUEST.'

:GOTO 10

30IF K%(0)>0 THEN 33

32PRINT 'THERE IS NO QUESTIONNAIRE STORED YET.' :GOTO 10

33FOR I=1 TO K%(0)

34PRINT N8S(I):NEXT I

:GOTO 10

6019%=INSTR(1,Z$,P45)

61Y%=RIGHT(Z$,[19]+4%

62GOSUB 1500:

63IF Z$=0% THEN 184

64IF LEN(Y$)>5 THEN Y%=LEFT(Y$,5%

73OPEN 'NFILE' AS FILE 8:DIM #8,C3S(0):C3S(0)=Y$;CLOSE 8;CLOSE 9

80K$(0)=2

81I=INSTR(1,Z$,P65)

82Y%=RIGHT(Z$,[19]+7%

83GOSUB 1500:

84IF Z$=0% THEN 137

85PRINT 'THERE IS NO SUCH QUESTIONNAIRE STORED YET. IT MIGHT BE'

86PRINT 'HELPFUL TO USE THE REQUEST CATALOGUE.'

87IF LEN(Y$)>5 THEN Y%=LEFT(Y$,5%

88KILLY$=KILLY$+'L':KILLY$+ '1':KILLY$+ 'Z':KILLY$+ '3'

89K%(0)=K%(0)+1

90FOR I=Z$ TO K%(0)

91N8S(I)=N8S(I+1)

92NEXT I

93GOTO 10

20019%=INSTR(1,Z$,P15)

201Y%=RIGHT(Z$,[19]+6%

202GOSUB 1500:

203IF Z$=0% THEN 208

205PRINT 'THIS QUESTIONNAIRE HAS ALREADY BEEN STORED. IT MIGHT BE'

206PRINT 'HELPFUL TO USE THE REQUEST CATALOGUE.' :GOTO 10
PRINT "HOW MANY QUESTIONS HAVE YOU GOT?"
INPUT R%:
PRINT "YOU CAN NOW TYPE IN THE NEW QUESTIONNAIRE. PLEASE TYPE : A2.2
PRINT "AT THE END OF EACH QUESTION AND *** AT THE END OF THE TEXT.
A2%=0: R2$="***": A3%=0: R3$="": B2$="": S1%=0: S2%=0
22 K%(0)=K%(0)+1
22 IF K%(0)>30 THEN K%(0)=30 ELSE 226
24 PRINT "SORRY BUT I CANNOT STORE MORE QUESTIONNAIRES."
GOTO 10
26 N8$(K%(0))=Y5
29 INPUT LINE Z 5
30 DIM A1$(100): A2%=A2%+1: A1$(A2%)=Z$: PRINT A1$(A2%)
33 IF INSTR(1: Z$: R2$)>0 THEN 236
34 GO TO 229
38 PRINT "NOW TYPE THE TRIPLETS OF NUMBERS, EACH TRIPLET PER LINE."
PRINT "THE NUMBERS SEPARATED BY COMMAS. PRINT 0, 0, 0 AT THE LAST ONE.
: OPEN Y$+1: AS FILE 4: DIM #4: T2%(0): N4%(40)
: OPEN Y$+2: AS FILE 2: DIM #2, N5%(40)
: FOR I=1 TO T2%(0)
: INPUT C%, D%, E%
: PRINT C%, D%, E%
: N4%(I)=C%: N5%(I)=D%: N6%(I)=E%
: NEXT I
: CLOSE 1: CLOSE 2: CLOSE 3
2600 OPEN $+1: AS FILE 4: DIM #4: T3%(0): L7%(40): T3%(0)=R%
: OPEN Y$ AS FILE 5: FIELD $5: S12% AS DS: FOR I=1 TO 5: FS=SPACE$(512%)
: LSETD$=FS: PUT #5: RECORD I: NEXT I
2610 FOR I=1 TO A2%: IF B1$=": THEN B1$=A1$(I)
ELSE B1$=B1$+A1$(I)
2680 IF INSTR(1: B1$: R3$)=0 THEN 320
2690 PRINT "269": B1$=
: I9%=INSTR(1: B1$: R3$): I3%=I9%-I15%=I9%+1
: B2$=LEFT(B1$, I3%)
: PRINT '273': B2$
: B1$=RIGHT(B1$, I5%)
: S1%=S1%+1: S2%=S2%+LEN(B2$)
: L7%(S1%)=LEN(B2$)
: PRINT '278': S1%, L7%(S1%)
: R%=S2%-(S5%/512%)*512%
270 PRINT 'S2%="': S2$=
250 K1%=S2%/512: IF R%>0 THEN K1%=K1%+1
251 M1%=R%: Z9%=0: FOR J=K1% TO K1%: GET $5: RECORD J: ES=DS: IF J>K% THEN 290
257 M3%=M%+1: IF K%<K1% THEN 289
258 IF E5=SPACE$(512)% THEN SSS=B2$: C
READY.
OLD MEDINB
READY
OLD MEDINA
READY
LIST 280-5000
MEDINA 03:40 PM 20-AUG-76
283 IF E5=SPACE$(512%): THEN SSS=B2$: ELSE SSS=LEFT(E5, M3%)+B2$: GO TO 293
2832%+512-M%+1:S5%=LEFT(E5, M3%)+LEFT(B2%, Z2%): 29%=Z9%+Z%: GO TO 293
```basic
90Z8%=Z%+1%;SS%=RIGHT(B2$,Z8%)
91PRINT'S2%=';S2%, 'B2$=';B2$, 'Z8%=';Z8%, 'Z%=';Z%, 'SS%=';SS$
92INPUT V%
93LSETDS=SS$;PUT #5,RECORD J;PRINT S5$;NEXT J
94IF R%=0 THEN K%=K%+1;ELSE K%=K%
96M%=R%+1
97:PRINT'297',K%,M%
98:IF S1%=T3%(0) THEN 315
100GOTO 268
115PRINT 'T3%(0)=';T3%(0);CLOSE4;CLOSE5
20NEXT I
25GOTO 10
30I9%=INSTR(1,ZS,P2$)
31Y%=RIGHT(ZS,I9%+6%)
32GOSUB 1500
33IF Z2%=0 THEN 184
340IF LEN(YS)>5 THEN YS=LEFT(YS,5%)
35OPEN 'ZFILE' AS FILE 7;DIM 7,C4$(0):C4$(0)=YS;CLOSE7;CLOSE9
CHAIN 'MEDINCS'
1500Z2%=0
1510T9%=LEN(YS)
1520:IF T9%<5 THEN T9%=5
1530FOR I=1 TO K%(0)
1540:IF LEFT(YS,T9%)=LEFT(NBS(I),T9%)THEN Z2%=I ELSE 1560
1550GOTO 1570
1560NEXT I
1570RETURN.
1570CLOSE 9
1571END
```

READY

BYE

CONNECT TIME: 9 MINUTES - 0 CU
CPU TIME: 11.5 SECONDS - 0 CU
DISK STORAGE: 59 BLOCKS SAVED
DEPOSIT 332 CU

JOB 1 USER 38.2 LOGGED OFF KB11 AT 20-AUG-76 03:42 PM
1 OTHER USER(S) STILL LOGGED IN UNDER THIS ACCOUNT.
GOOD AFTERnoon
LIST
MEDINDB 03:38 PM 20-AUG-76
10OPEN 'FILE' AS FILE 8: DIM # B$; C$(O)
: Y$= C$(O): CLOSE 8
: OPEN Y$ AS FILE 1
: FIELD #1,512% AS DS
: OPEN Y$+ 'L' AS FILE 3: DIM #3, T%(O); L%(40)
: OPEN Y$+ 'I' AS FILE 4: DIM #4, T%(O); N%(40)
: OPEN Y$+ '2' AS FILE 5: DIM #5, N2%(40)
810OPEN Y$+ '3' AS FILE 6: DIM #6, N3%(40)
: KS= I$ + 'B'= 0
: FOR I=1 TO T%(O)
: PX= 512%- M% + R% = L%(1)- P%: B%= B% + L%(1)
: IF R%> 0 THEN 90
86GET #1, RECORD KS
: ES= DS
: ES= MID(E$, M%, L%(1))
: PRINT ES: GO TO 100
88PRINT 'KS='JK%
90GET #1, RECORD KS: ES= DS: ES= RIGHT(E$, M%): R% = KS + 1
100GET #1, RECORD R$: ES= DS: ES= LEFT(E$, R%): ES= ES+ ES$: PRINT E$
102PRINT N1%(I), N2%(I), N3%(I)
: KS= B%/ 512%: KS= KS + 1
: M%= B%- (B%/ 512%) * 512% + 1
: NEXT I
: CLOSE3: CLOSE4: CLOSE5: CLOSE6: CLOSE 1
: CHAIN 'MEDINA.7'
: END

READY

OLD MEDING

READY

LIST
MEDING 03:40 PM 20-AUG-76
55GOSUB 1100
10OPEN 'FILE' AS FILE 7: DIM #7, C4$(O); Y$= C4$(O): CLOSE 7
30PRINT 'HELLO! I AM CONDUCTING A SURVEY ON LEISURE TIME AND HOBBIES.'
35PRINT 'AND I WOULD LIKE TO ASK YOU A FEW QUESTIONS ABOUT THESE.'
40PRINT 'IT WILL NOT TAKE MORE THAN TEN MINUTES AND IT WILL BE VERY
50PRINT 'HELPFUL TO US. WOULD YOU LIKE TO TAKE PART? PLEASE ANSWER'
60PRINT 'YES' OR 'NO'.
70INPUT Z$ 85GOSUB 1100
90IF Z$ = '7' THEN 100 ELSE IF Z$ = '8' THEN 90 ELSE IF Z$ = '9' THEN 90 ELSE 85
80PRINT 'SORRY BUT I CAN UNDERSTAND ONLY "YES", "NO", "?", ANSWERS.'
: GO TO 70
90PRINT 'ALRIGHT, THANKS FOR COMING ANYWAY. NOW PLEASE SEE THE SUPERVISOR.'
: GO TO 1000
100EI= TIME(0%): ES= 0
110PRINT 'THANKS VERY MUCH. NOW BEFORE STARTING I WOULD LIKE TO KNOW
120PRINT 'YOUR SEX. SO PLEASE TELL ME ARE YOU MALE? "YES" OR "NO"?'
130INPUT Z$ 115GOSUB 1100
117IF Z$ = '7' OR Z$ = '8' THEN 120
118PRINT 'SORRY BUT I CAN UNDERSTAND ONLY MEANINGFUL ANSWERS "YES", "NO"
119PRINT OR ?"." GO TO 115
120PRINT 'THANKS. I JUST WANT TO TELL YOU THAT THE QUESTIONS HAVE
125PRINT 'SIMPLE "YES" OR "NO" ANSWERS. SO YOU JUST PRESS THE "YES"'
125PRINT 'OR "NO" BUTTON IN FRONT OF YOU. IF FOR ANY REASON YOU DO
135PRINT 'NOT UNDERSTAND A QUESTION PLEASE PRESS THE "?" BUTTON.'
140PRINT 'FINALLY DO NOT FORGET TO PRESS THE "RETURN" BUTTON AFTER

A2.4
1·!,>OPIlINT·No~r

IF YOU FEEL.. READY PRESS THE "YES" BUTTON TO START.'
I 521 F Z$=' 7' THENR9 $=. ·1'lAL.E EL.SE . R9 S= • FE:1AL. E ~
155INPUT Z$
160IF 'ZS=' ,'TllEN' 180
1
.. ' ,
•
• \
',.
165PRINT 'YOU CAN TAI(E YOUR :TIME, BUT WHEN 'YOU . FEEL. . READY. PRESS' ,
I, :.
170PRINT'THS "YES'~ BUTTON ,.;O.START;··
..
,.
.., ....
,. . \ .
175GO TO 1 5 5 '
. '.
, '
.~. '. ,:'~.:l-. '
" ,., .
. 180 GOSUB I 100
. .
i· .
.'
,
' . ','
" "1,.., .
t . . , ,; t~~
;, . .
1
" .,
'I
': ":; \ ,;
, . , . l.'!r,:~ .
'" .~;.
635 OPEN Y $.+ '1.' 'AS FlI.E4: DIl1 '4, T %(0 n),.
::'! .
..'! "•• '
l J'
:OPEN Y$+'I' AS FII.E 1:.DI~#I,TI%CO),NI%C40)
.:.:
:.
•
:OPEN Y$+'2' AS FIL.E 2:DIt1u'2,N2?(40)·;
.'
.:·1
.. '. .. i
:OPEN YS+·3·. ASFII.E3:DII1I3',N3%(40)
"
".
,
'
. ,.
:OPEN Y$ A~ FIL.E5:FIE~D#5,512? ASD$iDI~U$(40)
,
.'
, .
.
I,
...•.
" . : . ."
; ,
6/j(JO .• =0
.
'666FOR J=I TO T%(O):U$CJ)"': ':NEXT J:K?=11'1%=IIQ%"IIA~~0:.·
'.:! ',"
675FOR J= I TO Q?:A%=A%+L.%CJ) :NEXT' J-.
.
. . ." . ' .
,
,
. 677P%=A%/512: R%=A%- 1'%*512: 1'.%=P%+ 11 R I %=R%
· ,,
678IF R1.>0· THEN 715 , ;
1I
.
. '/
, ;
.710P%=PlI-IIRU=512
.
1I
','
J'
715FOR I=K% TO P%:GET 15,RECORD I':E$=D$
.i
'1
720IFI>I(%' .THEN 780
.1'11;
. '.
\ 7351 F PlI>Kr. THEN 755
"i
,
,.'
737E$=1'lID(E$,Mr.,L.%(Q%»:PRINT E$:GO TO 860
,
:1\ I
755EIS=RIGHTCE$,M?) :GO TO 800' :'
.
.:
.: .
780E2S=L.EFT(E$,Rlr.)~~i"EI$+EisIPRINi E$'
,
'I ,:
.. )
. '.
800NEXT 1
; .\
. '11'
I'
" .'
I
. 'I
802 A%=O
'
I I . ! ! ... . '1, .
1303E2=Tli1EC0%)
t
:::'.
I
j.
1310INPUT Z$ .
11:
,:~,:
I"
B 1 1 GOSUB 1 10(3
.: . ",
. ;,' .
812E3=TII1EC0:0 IE4"E3-E2:E5=E5+E4
.'
:'.,:'
\
.
'
.,
13151F Z$"'7' THEN QI,.=NI%(Q'X) EL.SEIOF'Z$"·8·; THENQI%"N2r.CQ:O' .... : i.
t •
E L.SE IF ZS='9' THEN QIX=N3%(Q:;) E1.SE 825
.'
"
1320GO TO 830.
I.·
825PRINT'SORRY BUT' YOU CAN 'ANSWER ONL.Y :·YES ..... NO ... it ? ....
• I
·826GO TO 810 .
.
I
830U$(Q%)"Z$: IF QI%=12I T1tEN .9130 .
. .
I3 l I5Qr.=Qlr.
I' .
.
.
.' .'
847:·19:;=QI%.":.I:FOR J=I TO ,19r.::B?=O?+L..XCJ):NEXT JIRI,."B%IBr."(J .'
I:. ;.
855IC%=RIZ/512+I:MI=RIX-CRIX/512)*512+1:;:GO TO 675
'.
;
: 9(JOPRINT'THANICS FOR ANS\~ERING THOSE QUESTI0NS.'NO\~ PL.EASE 'SEE THE']
905PiuNT·SUPERVISOR.·
.
, ! " ,
I
910E6=TII1E(OX): E7=E6-EI
:.
. , ! ',
9150PEN "1(83:" FOR OUTPUT.AS FIL.E 7
917PRINT#7,Y$~R9$
920PRINTh7, ·TOTAL. TII1E='; E7, • TOTAl. REACTION .TIME.... ,; E51 CI.OSE7 "
925CL.OSEI :CL.OSE2: CL.OSE3ICLOSE4: CLOSES'
, .
,".
,.
iC00GHAIN ·:H:DINA··7
.,
.1100F9$=SPACE$CI:O .,
. .1:
.
.':
'.
.!
1150FORl=1 TO 251PRINT F9$:INEXT 1
/.\
'.,'
I
1180RETURN'
.
t
.
,'. . . . . i
".
)'
1200END
' .
(
. .'
'.
, .. ' "

l
I'"

,

I'

.

....

.

.

.

.

~(40)

t_

, ' . . '.

:,,;.

.

..

I' ,

.

", .

11

.,

\;

·

::!,'

i'

'

.

..

.

·..

!

I'

,

..

.'

.

i

.'_

READY

I
t'

I •

.

-'

p
d

I1
'I

.

1

i' .
'"

..

'r',

"

'J.

... .

. . . . \.
' .

1

.

.,

r

BYE
.'
:i
I,
"
~ONNECT TIMEi 9 MINUTES
- 0'CU
CPU Tlt1EI .7.3 SECONDS' - ·.0·:CU·.
DISK STORAGE: 53.BLOCKS SAVED
,
: DEPOSIT 332 CU
JOB 2 USE1i130.2 LOGGED OFF KB9 A1' 20-AUG'-76 03146P.'1
GOOD A F T E R N O O N i l

I,

..

.)

"

,"

,I

,

··1
.' .,'

.'


APPENDIX 3 - INTRODUCTION, E.P.s, C.P.s.
A TYPICAL INTRODUCTION AND ENDING OF A COMPUTER INTERVIEW

(INTRODUCTION)

1. HELLO. I AM CONDUCTING A SURVEY ON LEISURE TIME AND PEOPLE'S HOBBIES AND I WOULD LIKE TO ASK YOU A FEW QUESTIONS ABOUT THESE. EVERYTHING WILL BE KEPT CONFIDENTIAL. IT WON'T TAKE MORE THAN FIFTEEN MINUTES AND IT WILL BE VERY HELPFUL TO US. WOULD YOU LIKE TO TAKE PART? PLEASE ANSWER 'YES' OR 'NO'.

2. THANKS VERY MUCH. NOW BEFORE STARTING I WOULD LIKE TO KNOW YOUR SEX. SO PLEASE TELL ME, ARE YOU MALE? 'YES' OR 'NO'?

3. ALRIGHT THEN. THANKS FOR COMING ANYWAY.

4. I JUST WANT TO TELL YOU THAT THE QUESTIONS WILL HAVE SIMPLE 'YES' OR 'NO' ANSWERS. SO YOU JUST PUSH THE 'YES' OR 'NO' BUTTON AFTER EACH QUESTION. IF YOU WANT TO THINK ABOUT A QUESTION THEN PLEASE DO SO. IF FOR ANY REASON YOU DON'T UNDERSTAND A QUESTION PRESS THE '?' BUTTON. NOW IF YOU FEEL READY PRESS THE 'YES' BUTTON TO START.

   (MAIN QUESTIONS)

   

   (ENDING)

THAT IS THE END OF THIS SESSION. THANK YOU VERY MUCH FOR YOUR HELP. THE SUPERVISOR WILL NOW COME AND SEE YOU.
ENCOURAGING AND CHATTY PHRASES
The complete set of E.P.s and C.P.s derived from the four computer interview programs, analysed in Chapter 2, are given here for convenience of reference and for those who may wish to use them as a basis for further research and development as suggested in Chapter 13.

ENCOURAGING PHRASES

Pregnancy
'Thank you. You are managing to answer my questions very well' / 'Thank you' / 'We have almost finished' / 'Good' / 'Thank you' / 'We have almost finished' / 'Good' / 'Congratulations' / 'Good that's fine' / 'Fine' / 'Good I am pleased to hear that' / 'I am very sorry to hear that'.

Gastric Ulcer
'Thank you' / 'Well you are being very helpful so far. I hope you still find the interview interesting' / 'Thank you. We are getting on very well and I have only one or two more questions to ask you' / 'I see' / 'O.K.' / 'Very well'.

Sex
'O.K.' / 'I see' / 'Alright' / 'Very well' / 'Well that's all very clear' / 'Good' / 'I am so glad' / 'Well that's good' / 'Splendid' / 'Thanks for answering those questions' / 'Fine. I think that's clear' / 'Good. This is all very useful information and will help us to help you' / 'O.K. We've nearly finished now'.

Chest
'Fine' / 'You have been very helpful so far' / 'You have been very helpful and I've nearly finished' / 'I understand' / 'O.K.' / 'I see' / 'Thank you'.

A3.2
CHATTY PHRASES (PRECEDING THE QUESTION)

Pregnancy
'First of all' / 'For instance'.

Gastric Ulcer
'Well first of all' / 'Tell me', / 'First of all' / 'One
final question which I would like you to think carefully about' / 'Could you tell me' / 'Now I would like you to think carefully
about your answer to the next question'.

Sex
'Now first of all' / '(First) let me ask you' / 'Now the next
question' / 'Now finally' / 'Tell me', / 'Next' / 'Now I want
to ask you some more questions (of this kind)' / 'Firstly tell
me' / 'Now a few more questions' / 'Now here's a rather
important question' / 'Now' / 'Now I want to ask you' / 'Let me ask you just one or two more questions' / 'Firstly' / 'Here is another question' / 'This may be a difficult question
but' / 'Then, finally a very difficult question but do your best' / 'Firstly let me ask you' / 'Now finally' / 'First' / 'Incidentally' / 'Well now for some more (questions). Again do your best to
answer these truthfully'.

Chest
'By the way' / 'Tell me', / 'I'd like you to think carefully
about this now' / 'First of all' / 'Now' / 'Could you tell
me' / 'Now this is a very important question'.

CHATTY PHRASES (FOLLOWING THE QUESTION)

Pregnancy
'Press the 'YES' if it was either of these hospitals'.

Gastric Ulcer
'Push 'YES' or 'NO'.
'Please push 'YES' or 'NO'.
'Push 'YES' if you know where the navel is and 'NO' if you don't know'.
'Think carefully and push 'YES' or 'NO'.
'Push 'YES' if it does and 'NO' if it doesn't'.
'Press 'YES' or 'NO'.
'Push one of the buttons'.
'Push 'YES' if you have and 'NO' if not'.
'Press one button'.
'Press 'YES' or 'NO' please'.
'Press 'YES' if it does, and 'NO' if not'.
'YES' or 'NO'.
'If you do push the 'YES' button, if not push the 'NO' button'.
'Press 'YES' if you have and 'NO' if not'.

Sex
'Push one of the buttons'.
'Push 'YES' or 'NO'.
'Push 'YES' or 'NO'.
'Push 'YES' or 'NO'.

Chest
'Press one of the buttons now'.
'Push one of the buttons now'.
'Push one of the buttons'.
'If you're not quite sure what I mean push the '? button'.
'Answer 'YES' or 'NO' (if you don't normally need to clear your chest at all, you should answer 'YES').
'Push 'YES' if you have, 'NO' if you haven't and '?' if you are not sure'.
