

Conceptualizing in Assembly Tasks

Patricia Baggett
Psychology Department
University of Colorado

and

Andrzej Ehrenfeucht
Computer Science Department
University of Colorado

Conceptualizing in Assembly Tasks

Abstract

This paper gives a method to determine a person's hypothetical conceptualization of an object -- its breakdown into subassemblies, subsubassemblies, and so on -- from the person's sequence of requests for pieces used in constructing it. A technique is given to determine whether, given a group of conceptualizations, there is a typical one. The hypothesis that assembly instructions presenting a typical conceptualization will yield better structural and functional performance than those presenting a minority one is supported experimentally. Conceptualizations are derived from objects built from memory (and incorrectly) by people who first studied typical or minority instructions. A new distance measure determines how far these conceptualizations are from those presented in the instructions. People studying typical instructions yield typical conceptualizations, and importantly, people studying minority instructions also yield typical conceptualizations, although they are significantly less typical than those from typical instructions. From the theoretical construct of conceptualizations and the methods of measuring them a practical principle, and a way to implement it, are found. The principle: When a single set of procedural instructions is designed, it should present the conceptualization that the majority of people to be instructed by it bring to the situation naturally.

Conceptualizing in Assembly Tasks

The statement that behavior is based on an underlying mental organization or conceptualization is not controversial. But a major problem is to infer the conceptualization from behavior, and to show that the conceptualization is psychologically valid. This study gives a method to derive a person's hypothetical conceptualization of an object he or she is building from a sequence of behaviors--requests for pieces used in constructing the object.

It also gives a method to determine, from a group of sequences from different people, or from the same person on different trials, how similar the conceptualizations are. The analysis allows us to determine, from a group of conceptualizations, whether there is a "typical" one. We say what it means for a conceptualization to be presented in instructions. We test the hypothesis that instructions which present a typical conceptualization will yield better performance than those presenting a minority conceptualization. Finally, we analyze the conceptualizations derived from objects built from memory by people who first study either typical or minority instructions. The question in this last analysis is, when a particular conceptualization is studied, how well will that conceptualization still be manifested in the object built later from memory?

Research on inducing conceptual structure from a person's sequential output or from stimulus input combined with subject output is long and varied. Reitman and Rueter (1980) presented a review of several studies and some techniques used. We give only a brief summary here. Bousfield (1953), Bushke (1976), and Reitman and Rueter (1980) studied clustering in recall as a measure of mental organization. Chunking, including temporal parameters, in Go (Reitman, 1976), chess (Chase & Simon, 1973), and electronics (Egan & Schwartz, 1979) has been viewed as indicating what units are stored in memory. Pauses in verbal input (Wickelgren, 1967; Bower & Winzenz, 1969) and output (Chafe, 1977; Bower & Springston, 1970; McLean & Gregg, 1967) have been used to infer memory structure.

Our method also uses sequence information, namely, order of request of pieces in an assembly task. Our underlying hypothesis is as follows. When a person is asked to build an object, from memory or using a model as a guide, the person conceptualizes the object as consisting of subassemblies, subsubassemblies, and so on, and groups his or her requests for pieces according to the conceptual division. This notion is not new. For example, the research, using verbal material, of Monk (1976) and Reitman & Rueter (1980) is based on the idea that subjects chunk the material and recall all elements of one chunk before proceeding to the next.

To our knowledge, our research is new in that: (1) it describes in precise detail how to derive a hypothetical conceptualization from sequence information in an assembly task, which is primarily nonverbal; (2) it gives a method to determine whether different conceptualizations are uniform; (3) it presents a test of the hypothesis that "typical" procedural instructions result in better performance than minority ones; and (4) it gives a technique for comparing the conceptualization presented in instructions to the conceptualization manifested in a subject's performance from memory

after studying the instructions. As mentioned above, the question this analysis can answer is, how close in conceptualization is what goes in to what comes out?

The Technique

Here are the experimental methods for finding an individual person's conceptualization, for determining whether subjects conceptualize uniformly, and for choosing the most typical conceptualization for a population of subjects. The methods require extensive programming. (See Perry, Note 4.) Variants of some of these methods are given in Baggett (1983) and Baggett and Perrig-Chiello (Note 3).

1. Finding the natural conceptualization of an individual.

We give here a technique to find how people divide an object into subassemblies (how they conceptualize it) from the order in which they request the parts in constructing the object. The assumption we are making can be illustrated by a simple example. If, in selecting four pieces, A, B, C, and D, to build a column A-B-C-D, the person requests A and B, and then D and C, we expect that in a division into two parts, the person has the concepts AB and CD.

The method used is to have a person ask for pieces one at a time and to record the order of request. (If two or more pieces in the object are identical, they must be made distinct in some way--for example, marked with numbers--so that the exact order of request of each piece in the final product can be determined.) The data are easy to gather, even for complex objects. We have data from an object (the one shown in Figure 1)

Insert Figure 1 about here

consisting of 80 pieces, but we think that substantially more pieces can be easily handled. The data analysis is straightforward. It consists of three parts:

a. An object for assembly is represented by an abstract graph whose nodes represent pieces and whose edges (links) represent connections. (This representation can be used on any assembled object, not just pieces from the Fischer-Technik assembly kit used here.) The graph of the object in Figure 1 is given in Figure 2. Nodes in Figure 2 are numbered 1 to 80 to correspond to specific pieces in the object.

Insert Figure 2 about here

b. A distance between connected nodes on the graph is introduced, based on how close together the requests for the two pieces are. For example, if a person requests piece 1 fifth and 4 second, the distance between pieces 1 and 4 is $|5-2| = 3$.

c. A cluster analysis is performed, and the clusters are used as hypothetical conceptual units of the assembler. The method of clustering that we use is as follows. Each node is put in a cluster with its closest connected neighbor. An example is given in Figure 3 by the dotted lines in the figure. Then each cluster is put in a higher-order cluster with its closest connected neighbor. These are the solid lines in the figure. This process is continued until all clusters fall into the same higher-order cluster. The analysis was done on a VAX 11/780 under the UNIX operating system, using a programming package written by Perry (Note 4).

The method yields a hierarchical tree, which is the hypothetical natural conceptualization of the object by an individual.

2. Finding whether different people conceptualize uniformly.

Here we give a method to determine how different are the conceptualizations of different people, and of one person on different trials. Are they minor variants of the same conceptualization, or do they form different categories?

a. We first restrict ourselves to comparing conceptualizations in which the final products built are identical. This means that the underlying abstract graphs (as in Figure 2) and the actual pieces used are the same. For example, when adults use a physical model as a guide in building, they typically copy it perfectly and thus construct identical objects.

As in (1) above, the subject is required to request each piece separately, and the sequence of requests is recorded. A person's conceptualization of the object is derived from the order of request using a computer program (Perry, Note 4).

Each conceptualization (tree) becomes a point in a space in which its distance from every other point is calculated. The distance between conceptualizations is described in Appendix 1. A cluster analysis is performed on the space of points. (This cluster analysis is the same as that given in (1) above.) Those points occurring in the same first-order cluster are assumed to be variants of the same conceptualization. If a majority of conceptualizations fall into one first-order cluster, we assume that most subjects conceptualize uniformly. We then define the most typical conceptualization as that conceptualization in the majority cluster which has the smallest average distance to all other conceptualizations in its cluster.

b. Now suppose we want to compare conceptualizations in which the final products built are different, that is, the trees have different leaves. (This will usually be the case, for example, when subjects are building a complex object from memory. They don't do it perfectly, and each person's product is unique.)

The method is the same as above, but the distance between conceptualizations is different. It is given in Appendix 2, together with a complete simple example.

Using the Technique

In the experiments reported below, we attempted to determine whether our method of measuring conceptualization was reasonable. In Experiment I, we derived conceptualizations from 47 trials of subjects building the object in Figure 1 from a model, and we determined whether conceptualization was uniform. We also derived the most typical conceptualization and a minority conceptualization maximally distant from it. In Experiment II we prepared two videotapes, one showing assembly of the object using the typical conceptualization and the other showing the minority conceptualization. Subjects watched one or the other videotape and then built the object from memory. The hypothesis was that performance from memory should be better for those viewing the typical tape, because it matches their "natural" conceptualization. In extended data analysis in Experiment II, we examined conceptualizations in the objects built from memory from those who watched the typical versus the minority tape. We assessed the subject's conceptualization, as a function of the conceptualization presented in the original instructions, to discover how well the presented conceptualization was manifested in a person's performance.

EXPERIMENT I

Method

Subjects. Twenty-one students (11 female, 10 male) from the Psychology 100 subject pool at the University of Colorado participated as part of a course requirement. One participated in a one-hour session, fourteen in two one-hour sessions (48 hr apart) and six in three one-hour sessions (also 48 hr apart).

Materials. Fischer-Technik 50 assembly kits were used. Each kit contains 48 different and 120 total red, grey, silver, and black plastic and metal pieces. The largest is 90 x 45mm (3.54in x 1.77in) and the smallest 5mm² (.2in²). The model shown in Figure 1 was the object to be built. It consists of 80 pieces and 104 physical connections.

Procedure. Subjects were tested for either one, two, or three sessions. They were told that in each session they would build an object from an assembly kit, using a model as a guide. In session 1, they filled out a questionnaire indicating sex, major, and familiarity with Fischer-Technik and with Lego and other building kits. The experimenter explained that they would ask for each piece, by name or by pointing, as they built. To familiarize them with the names beforehand, subjects first did a matching task. A box with one of each of the 48 different pieces in the kit, and sheets with the 48 names, were placed before them. They were instructed to place each piece by its correct name. They were told that this was not a test, and if they were not sure about a match, they could ask the experimenter. (The names, selected using a schema given in Baggett and Ehrenfeucht (1982) and Baggett (1983), were short and easily matched (96% correctly in a previous experiment) with their physical referents.) The experimenter checked the matches and corrected errors. Pieces were placed back in the box.

Subjects were then given a folder containing color photos of the pieces and their names, and the box of pieces. The object to be built was brought

into view and set before them. Subjects were allowed to handle it, make it function, and disassemble parts if they wanted. They built the object, asking for pieces one at a time as they went. They used the folder and the box of pieces in deciding which pieces they wanted. The experimenter recorded the sequence of requests, and handed the subject the requested pieces.

Subjects returning for sessions two and three built the object a second and third time, using an identical procedure but with no matching task.

Results and Discussion

The 47 orders of request made by people building the object in Figure 1 were analyzed using the cluster analysis program of Perry (Note 4). As described in Appendix 1, each conceptualization (tree) became a point in a space, and its distance from every other tree was computed. A cluster analysis was performed on the space of 47 points. Six first-order clusters were found. The largest contained 33 trees, or 70%, and the remaining five contained 3, 3, 3, 3, and 2 respectively. We interpreted the result to mean that 70% of the conceptualizations were minor variants of one another. There were five other different conceptualizations, which were not variants of those in the majority cluster. The average session number for the 33 trees in the majority cluster was 1.82, while the average session number for the 14 others was 1.35. Thus in later sessions people tended to become more uniform.

We define the most typical conceptualization as that one among the majority group of conceptualizations which had the smallest average distance from the others in its group. The most typical conceptualization for the majority group came in session 2 and was from a female subject. She was a nonscience major who rated her experience with Lego and similar assembly kits as one on a scale of 0 to 3. The conceptualization is shown in Figure 2. It will be referred to in the rest of this paper as the typical conceptualization.

Insert Figures 3 and 4 about here

We selected the conceptualization from the remaining 14 which was maximally distant from the typical one. It is shown in Figure 4 and will be referred to as the minority conceptualization. This conceptualization was given in session 1 by a male engineering major who rated his familiarity with assembly kits as 2 on a 0 to 3 scale. It came from the smallest cluster, with only two trees.

The result of Experiment I agrees with that found in Baggett, 1983. In the earlier study, about 80% of trials assembling an object different from the one used here (54 pieces, 58 connections) clustered together. If these findings of a large uniformity in conceptualization generalize, this means that in individualized instruction, just one set of instructions can cover a large majority of subjects.

EXPERIMENT II

The goal of Experiment II was to determine how performance would vary as a function of the conceptualization (typical vs. minority) presented in instructions. The theoretical hypothesis was that typical should be better than minority.

We wanted the instructions to be in the form of narrated videotapes. Our first concern was how to find good names for the object, its subassemblies, and subsubassemblies, as indicated in Figures 3 and 4, to be used in the narration. This required an experiment, IIa.

Experiment IIa

Finding Names for Subassemblies

Method

Subjects. Twenty four subjects from the Psychology 100 subject pool at the University of Colorado participated as part of a course requirement. They were randomly divided into two 12-member groups.

Materials. The physical object shown in Figure 1 was used, together with its physical breakdown into subassemblies (the solid lines in Figures 3 and 4) and its further physical breakdown into subsubassemblies (the dotted lines in Figures 3 and 4).

Procedure. Subjects were assigned to one of two groups. Group one built the object according to the conceptual breakdown in Figure 3 (the typical one), and then named the 3 subassemblies, the 11 subsubassemblies, and the whole object. Group two did the same for the minority conceptualization, containing 4 subassemblies and 14 subsubassemblies.

As in Experiment I, subjects first matched pieces with names, and they built by requesting pieces one by one. They were instructed that they had to build the object broken down into the subassemblies lying before them, and that they could build the various subassemblies in any order. The experimenter recorded the order of request, to enforce the conceptual subdivision.

After building, they were asked to name the whole object, each subassembly, and each subsubassembly. It was explained that the best names would be selected and used in the narration to our planned instructional videotape. An answer sheet was provided. It contained blanks with numbers corresponding to numbers on each of the physical subassemblies. The sessions lasted approximately one hour.

Results and Discussion

Using the naming schema given in Baggett and Ehrenfeucht (1982) and improved in Baggett (1983) we selected the most frequently generated name for the object and for each subassembly. (This same schema had been used earlier to derive names for pieces in the kit.) The names are given in Figure 5 for the typical conceptualization and in Figure 6 for the minority

Insert Figures 5 and 6 about here

one. The average length of a name for the 15 nodes in the typical tree is 2.0; for the minority tree (with 18 nodes) it is 2.16. The object's name, based on the naming schema, is lift.

Experiment IIB.

How the conceptualization in instructions influences
conceptualization and performance in assembly

Method

Subjects. Fifty-eight subjects from the Psychology 100 subject pool at the University of Colorado participated as part of a course requirement. They were divided into two 29-member groups, each with 13 females and 16 males. Group assignment was as random as possible, with the constraint that groups contain particular numbers of males and females.

Materials. Two narrated videotapes showing assembly of the lift were prepared. They were shot and edited by James Otis, a professional filmmaker at the University of Colorado. Mr. Otis was given the two conceptualizations shown in Figures 3 and 4, together with the actual orders of request which gave rise to the conceptualizations. He prepared two videotapes of approximately equal length (23 min for typical, 22 min for minority). His constraints were that the conceptualizations presented should be those he was given, and the order of assembly in each case should follow the order of request of the original subject whose conceptualization he was given. The 3/4 in videotapes were narrated by the author. Names for subassemblies came from Experiment IIa. The typical videotape contained 1598 words; the minority contained 1483. When shots were similar in the two tapes, as much as possible the narration was kept identical.

The videotapes presented the conceptualizations top-down breadth first. That is, first the lift was shown (and a demonstration of how it functions was given). Then, in the typical case, the three subassemblies were shown and assembled into a lift. Then the five subsubassemblies of the first subassembly were shown and assembled into the first subassembly, etc. Finally the individual pieces of the first subsubassembly were put together to make it, etc. The minority conceptualization was filmed similarly. The actress for both videotapes was Cynthia Russell. Only her hands and arms were shown. The shots were over her shoulder, and the background was a constant medium blue.

Procedure. Subjects were run individually in a single session lasting from just over an hour to 2-1/2 hours. They first filled out a questionnaire identical to the one used in Experiment I. They were told that they would watch a videotape showing assembly of an object from an assembly kit, and that afterwards they would be asked to build the object from memory. They were further told that the videotape was narrated, so that pieces would be named. To give them a headstart on recognizing pieces and their names, they were told, they would first do the same matching task required in Experiments I and IIa. The experimenter checked their matches

as before and corrected errors. The experimenter explained that not every piece in the kit would occur in the object, and some pieces would occur more than once.

Subjects then positioned their chairs in front of a 15" color television set at a distance they selected, and then the appropriate videotape (typical or minority) was shown. Subjects had the option for the room's light to be on or off. (The room was without windows.) And they could adjust the volume to their preferred level.

When the tape was finished, the experimenter turned off the equipment and told the subject about requesting pieces one at a time (as in the previous two experiments). As before, a folder with pieces' photos and their names was beside the subject, together with a collection of one of each of the 48 different pieces. (The lift contains 24 different and 80 total pieces. Thus the subject could choose pieces for the memory trial that did not occur in the actual lift.)

The experimenter explained that there was no time limit, that the subject, after asking for a piece, did not have to use it, and that the objective was to build a lift as similar to the one in the videotape as possible. The time the first piece was requested was recorded, together with the time the subject indicated he or she was ready to stop. The experimenter recorded the order of request of each piece.

Results and Discussion

Measuring Performance in the Products Build from Memory

We first assessed how similar the lifts built from memory were to the original lift. One measure we used was based on structure. Namely, how similar in structure was a memorial lift to the original? This was operationalized as follows. An abstract graph of each lift was drawn. As in Figure 2, it contained nodes and links, indicating the pieces present and how they were physically connected. We counted the number of correct connections in the lift (104 were possible) and used this as our structural measure. Subjects were not penalized for incorrect connections, and correct connections could be wrongly oriented and still receive full credit. Our second measure was functionality. Namely, did the lift built from memory contain a handle which, if turned, caused a lifting device to travel up and down a tower? Subjects' lifts were given either full credit or no credit for functionality. Table 1 presents the structural and functional results, divided by gender of subject and videotape viewed.

Table 1

Average number of correct connections in lifts built from memory (104 possible) and number functional (in parentheses), as a function of gender of subject and videotape viewed.

		Videotape Viewed	
		Minority	Typical
gender	female	27.4 (0 of 13 functional)	37.2 (4 of 13 functional)
	male	42.3 (10 of 16 functional)	52.6 (15 of 16 functional)
	weighted mean	35.6 (10 of 29 functional)	45.7 (19 of 29 functional)

Note: There were 16 males and 13 females in each group.

A 2x2 between subjects ANOVA was performed on the structural and functional scores. For structure, there was a main effect of gender ($F(1,54)=11.6, p<.01$) and a main effect of videotape viewed ($F(1,54)=5.13, p<.05$), and no interaction ($F<1$). ($MS_{error}=283.9$.) For functionality, the same two main effects held, again with no interaction. (For gender, $F(1,54)=40.9$; for videotape viewed, $F(1,54)=10.0$; and for gender x videotape, $F<1$. $MS_{error}=.14$.)

Time to work during the memory trial and total number of pieces requested were also analyzed, each in a 2x2 ANOVA. Table 2 presents the means, again by gender and by videotape viewed. For time to work, neither of the main effects nor the interaction reached significance. For number of pieces requested, males requested significantly more than females ($F(1,54)=10.3, p<.01, MS_{error}=212.4$), while there were no other significant effects.

Table 2

Average time to work during the memory trial (in min) and number of pieces requested (in parentheses) (there are 80 in the original lift), as a function of gender of subject and videotape viewed.

		Videotape Viewed	
		Minority	Typical
gender	female	65.4 min (62.3 pieces)	64.6 min (65.5 pieces)
	male	55.4 min (73.9 pieces)	59.4 min (78.6 pieces)
	weighted mean	59.9 min (68.7 pieces)	61.8 min (72.7 pieces)

The results are straightforward to interpret. First, males outperformed females on three of four measures in assembly from memory: structure, functionality, and number of pieces requested. The effects of most interest in the study were the main effects of videotape viewed, regardless of gender, for structure and functionality. Namely, subjects in general perform better in terms of both similarity of structure and functionality when they first view a videotape presenting a typical, rather than a minority conceptualization of the object to be assembled. And this better performance is not accompanied by a longer time to work; this was not significantly different for those viewing the typical vs. the minority videotape.

On questionnaires given at the beginning of the experimental session, subjects were asked to rate how much they had played with Lego, erector sets, or other assembly kits, using a scale of 0 (never), 1 (once or twice), 2 (fairly frequently), and 3 (lots). The means of these subjective ratings are given in Table 3. We note that the ratios of the male to female familiarity ratings ($1.75/1.08=1.62$ for minority; $1.69/1.15=1.47$ for typical) are similar to the ratios of the male-to-female structural scores ($42.3/27.4=1.54$ for minority; $52.6/37.2=1.41$ for typical). Thus, one interpretation of the large gender differences in performance is that they arise from a difference in previous experience with similar tasks. In a later study (Baggett, Note 2) we will show an instructional design change that causes the gender difference to disappear.

Table 3

Mean rating of how much subjects had played with assembly kits, on a scale of 0 (none) to 3 (lots).

	videotape viewed	
	minority	typical
female	1.08	1.15
male	1.75	1.69

That typical instructions give better performance than minority ones is not really surprising. The typical conceptualization was defined as the one with the smallest average distance from those in a majority cluster containing 70% of the conceptualizations. Thus it was closest to the way most people conceptualized the task naturally. Let us assume that most of those instructed prepared their conceptual divisions by breaking down the lift into parts and subparts in a particular way at the beginning of the videotape (as the object's functionality was demonstrated). Then when the actual breakdown was presented, those whose already-prepared structure matched it had less to learn and no interference from two different potentially competing structures, the memorial one and the actual one.

One indication that the typical conceptualization better matched that of the majority of the subjects tested comes from the standard deviations in

the structural scores. For the typical videotape, the female s.d. = 9.3, and the male s.d. = 18.2. For the minority, the s.d.s. were 12.5 and 20.7 respectively. This means there was a larger range of scores for minority than typical instructions. For an occasional subject, perhaps the minority instructions fit his or her natural conceptualization, so that performance was especially high, and such scores inflated the minority standard deviations.

We do not know how stable conceptualization, as measured here, is. At this point it is an empirical but unanswered question.

Measuring Conceptualization in the Products built from Memory

The following question now arises. People built lifts from memory after viewing one of two videotapes. Each lift was completely unique. The order in which it was built, and the final graph, indicating the pieces used and how they were connected, was constructed. Can we determine what conceptualization was manifested in each lift built from memory? In particular, we want to know about conceptualizations from those viewing the typical versus the minority videotape. Are they closer to the conceptualization presented in the typical or the minority videotape?

The question is interesting because, if we can answer it, we will be able to assess whether people manifested the conceptualization given in their instructions, even if it did not match their natural one.

The analysis required is a variant of the analysis used above. Every lift built from memory, and the lifts built in the two videotapes (called the perfect typical and the perfect minority lifts) represent trees. Every memorial lift's distance from the perfect typical and the perfect minority is calculated. The question is whether the lifts built by the people viewing the typical videotape are closer to the typical than the minority conceptualization, and similarly, whether the lifts built by those viewing the minority videotape are closer to the minority than the typical conceptualization.

A new problem here is how to compute distances between trees whose end leaves do not consist of the same pieces. The distance that we use is given in Appendix 2. Again, the programming package of Perry (Note 4) was used.

We computed the following distances:

- (1) distance (perfect typical, perfect minority) = 321.
- (2) distance (conceptualization of each of 58 subjects, perfect typical)
- (3) distance (conceptualization of each of 58 subjects, perfect minority)

Thus we knew, for each subject, his or her distance from the perfect typical and minority. We rescaled the line from minority to typical to run from -100 to 100. We project each subject's conceptualization down on this line and ask where it hits. We call the score an x-score. The schema is illustrated in Figure 7. The predictions of course are that, if viewers

 Insert Figure 7 about here

follow their instructions perfectly, then the mean x-score of the people viewing the typical videotape ought to be about 100, and the mean x-score of those viewing the minority videotape ought to be about -100.

The results are in Table 4. A 2x2 between groups ANOVA (gender x videotape) on the x-scores yielded only one significant effect, a main effect of videotape viewed, $F(1,54)=5.30$, $p<.05$, $MS_{error} = 2792$. The x-scores of those viewing the typical videotape are not significantly less than 100 by a one-sample t-test ($t(29df)=2.04$). And those of people viewing the minority videotape have two important properties. First, they are significantly less than those of people viewing the typical videotape ($t(56df)=2.48$, $p<.02$). Second, they are far removed from the x-score of the minority videotape (statistics not necessary; -100 is significantly less than 51.2).

Table 4

Mean x-score as a function of videotape viewed and gender of subject.

	videotape viewed	
	minority	typical
female	60.6	85.0
male	43.6	83.5
weighted mean	51.2	84.2

Note. The x-score of the perfect typical lift was 100; the x-score of the perfect minority lift was -100.

The results of the conceptualization analysis tell us that typical instructions reinforce typical performance. Minority instructions do not reinforce a minority conceptualization. We do not know the reason for this. We do not know if such instructions are ignored, or whether they compete with a person's natural conceptualization. In the former case, a subject might think that the minority instructions are unnatural, and that his or her best strategy might be to block them out as much as possible and put together the lift in a natural way. In the latter case the subject might try to learn the minority conceptualization and find that, during the memory trial, he or she is actually working with two competing conceptualizations. Our experiment cannot decide between these alternatives.

Our major practical result is that performance on an assembly task is better when the conceptualization presented in instructions agrees with the one that people naturally bring to the task. As already mentioned, the result is not surprising. The main value of the work is theoretical and methodological. Namely:

- a. We specify precisely what it means to say that a set of instructions contains a conceptualization.
- b. We specify how to find out how different people, given an object to build, conceptualize the object.
- c. We specify how to determine if people in (b) who build the same object have basically the same conceptualizations (variants of each other) or substantially different ones.
- d. We specify what it means to say that an object has a natural, or typical, conceptualization.
- e. We can measure how conceptualization in assembly, even on imperfectly built objects, varies with the conceptualization presented in instructions.
- f. We give a blueprint for designing instructions so that the natural conceptualization is presented in them.

There are other ways that instructions might be improved. For example, individualized instructions might best be designed to match a specific person's idiosyncratic conceptualization. Or it might be that there are several different typical conceptualizations (for example, engineers' conceptualizations might be substantially different from novices'), and people should be instructed, based on subject variables, via those typical instructions that are judged best. Other improvements in videotaped instructions might come through changing narration, or through changing shot sequences based on ideas of visual cohesion and short term memory load (see Baggett, Note 2).

We hope that the methodology, theoretical approach, and practical results presented here will be useful to researchers and practitioners in a variety of situations involving sequences of behaviors. Our approach is currently being extended (Baggett, Note 1) to include actions and other elements, in the more general context of any procedural task.

Footnote

This research was supported by Office of Naval Research contracts #N00014-78-C-0433, N00014-84-C-0112, and N00014-85-K-0060. We wish to thank James Otis for shooting the videotapes used as stimulus materials, R.M. Perry for his programming package used in the data analysis, and Gabriela Jaramillo, Ken Johnson, Sharon Kaplan, Paul Perales, Cynthia Russell, Pete Tegner, and Lynn Wallick, student assistants who helped run subjects and prepare data for analysis. Some of the results were reported at the twenty-fourth and twenty-fifth annual meetings of the Psychonomics Society, 1983 and 1984. This work is Technical Report No. 139 of the Institute of Cognitive Science.

Reference Notes

Baggett, P. A multimedia knowledge representation for an "intelligent" computerized tutor. Talk given at ONR contractors' meeting on Instructional Technology for Advanced Training Systems, Atlanta, January 1985.

Baggett, P. How sequencing and visual cohesion in videotaped procedural instructions influence performance. In progress, 1985.

Baggett, P., & Perrig-Chiello, P. A method to analyze how children and adults conceptualize and build in an assembly task. Manuscript submitted for publication, 1985.

Perry, R.M. Computer Techniques for Cluster Analysis. Technical Report, Computer Science Department, University of Colorado, Boulder, CO, 1983.

References

- Baggett, P. Four principles for designing instructions. IEEE Transactions on Professional Communication, 1983, PC-26(3), 99-105.
- Baggett, P., & Ehrenfeucht, A. How an unfamiliar thing should be called. Journal of Psycholinguistic Research, 1982, 11, 437-445.
- Bousfield, W. The occurrence of clustering in recall of randomly arranged associates. Journal of General Psychology, 1953, 30, 149-165.
- Bower, G., & Winzenz, D. Group structure, coding, and memory for digit series. Journal of Experimental Psychology, 1969, 80(2, pt. 2).
- Bushke, H. Learning is organized by chunking. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 313-324.
- Chafe, W. Creativity in verbalization and its implications for the nature of stored knowledge. In R. Freedle (Ed.), Discourse Production and Comprehension, Vol. 1. New Jersey: Ablex, 1977. Pp. 41-55.
- Chase, W., & Simon, H. The mind's eye in chess. In W. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973. Pp. 215-281.
- Egan, D., & Schwartz, B. Chunking in recall of symbolic drawings. Memory and Cognition, 1979, 7(2), 149-158.
- McLean, R., & Gregg, L. Effects of induced chunking on temporal aspects of serial recitation. Journal of Experimental Psychology, 1967, 74, 455-459.
- Monk, A. A new approach to the characterization of sequential structure in multi-trial free recall using hierarchical grouping analysis. British Journal of Mathematical and Statistical Psychology, 1976, 29, 1-18.
- Reitman, J. Skilled perception in Go: Deducing memory structures from inter-response times. Cognitive Psychology, 1976, 8, 336-356.
- Reitman, J., & Rueter, H. Organization revealed by recall orders and confirmed by pauses. Cognitive Psychology, 1980, 12, 554-581.
- Wickelgren, W. Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. Quarterly Journal of Experimental Psychology, 1967, 19, 97-102.

Appendix 1

There are two steps in the cluster analysis for a group of conceptualizations which we used in Experiment I. Both use the computer program of Perry (Note 4).

1. Find the distance between all pairs of conceptualizations. The distance between conceptualizations on trials T_i and T_j is the sum (over all 104 connected pairs of pieces in the lift) of the differences in height in a conceptualization necessary to put a connected pair in the same cluster. Here is an example. Consider a pair of connected pieces p_1 and p_2 . Suppose they are placed in the conceptualizations T_i and T_j as shown in Figure 8.

Insert Figure 8 about here

In conceptualization T_i , p_1 and p_2 are in the same first-order cluster; the height is 1. In conceptualization T_j , p_1 and p_2 are in the same second-order cluster (dotted); the height j is 2. The distance between the pair of pieces (p_1, p_2) in conceptualizations T_i and T_j is the difference in heights, $2-1 = 1$. The distance between T_i and T_j is the sum (over all 104 pairs) of these distances.

2. Do a cluster analysis on the space of all pairs of conceptualizations with distances defined from step 1. A cluster analysis is done on the conceptualizations with each one put in a cluster with its closest neighbor.

Appendix 2

This section gives the technique for determining how different conceptualizations (hierarchical trees) from different people are. Each tree becomes a point in a space, and its distance from every other point is calculated. A cluster analysis (the one given in Appendix 1) is performed on the points in the space.

Let us take two conceptualizations X and Y and see how their distance apart is calculated. The method applies even if the graphs and the pieces used in X and Y are different.

X consists of a set, some subsets, some subsubsets, etc. We first make a list of all of these sets as follows. Each set contains each of 48 different pieces with some multiplicity. Each set can thus be described as a vector with 48 co-ordinates. Suppose x_i , a subset of X, contains one piece number 3 and two pieces number 4. (See Figure 4 for the meaning of the piece numbers.) Then x_i can be represented as (0,0,1,2,0,...0,0). Such a representation is given for all subsets of X, and for all subsets of Y.

For each set x_i in X, we find a set in Y that is minimally distant from it, using the Euclidean l_1 norm (the sum of the absolute values of the differences of the co-ordinates). Similarly, for each set y_j in Y we find a set in X that is minimally distant from it. The distance from X to Y is:

$$\begin{aligned} \text{distance } (X,Y) &= \sum_{\substack{\text{all sets} \\ x_i \text{ in } X}} \min (\text{distance } (x_i, \text{each } y_j)) \\ &+ \\ &\sum_{\substack{\text{all sets} \\ y_j \text{ in } Y}} \min (\text{distance } (y_j, \text{each } x_i)). \end{aligned}$$

This metric is only one of many possible ones. We chose it because it allows us to find distances between conceptualizations of objects which are not identical.

In our analysis, we calculated the following distances

- (a) distance (perfect typical, perfect minority) = 321
 - (b) distance (conceptualization of each of 58 subjects, perfect typical)
 - (c) distance (conceptualizations of each of 58 subjects, perfect minority)
- (Our analysis was done using a programming package written by Perry (1983).)

To be more concrete about how the distance between trees is calculated, we give a complete simplified example. Suppose four different pieces, A, B, C, and D, are available for building. Suppose there are 2 pieces C available, and one each of A, B, and D. Person S selects as follows. First a piece C is taken. Piece A is taken second and joined to C. Piece B is

selected third. Piece C' (identical to C) is taken fourth and joined to B. Finally C' is joined to C. The tree structure of S can be represented as ((A) (C' B)).

Person S' selects piece D first. Second is piece A and third is C. When C is taken, it is connected to A. The C and D are connected. The tree structure of S' is ((D) (A C)). We want to find the distance between the two conceptualizations.

The conceptualizations of S and S' consist of sets, subsets, etc., as explained above. We first make a list of the sets in S and in S', and represent them as vectors:

S		S'	
elements in set	vector	elements in set	vector
A B C C'	(1, 1, 2, 0)	A C D	(1, 0, 1, 1)
A C	(1, 0, 1, 0)	A C	(1, 0, 1, 0)
B C'	(0, 1, 1, 0)	D	(0, 0, 0, 1)
A	(1, 0, 0, 0)	A	(1, 0, 0, 0)
B	(0, 1, 0, 0)	C	(0, 0, 1, 0)
C	(0, 0, 1, 0)		
C'	(0, 0, 1, 0)		

For each set in S and S' we find a set in the opposite conceptualization that is closest to it, and we record the distance apart of the 2 sets:

a set S (represented as a vector)	a set in S' closest to it	distance apart
(1, 1, 2, 0)	(1, 0, 1, 0)	2
(1, 0, 1, 0)	(1, 0, 1, 0)	0
(0, 1, 1, 0)	(0, 0, 1, 0)	1
(1, 0, 0, 0)	(1, 0, 0, 0)	0
(0, 1, 0, 0)	(0, 0, 0, 1)	2
(0, 0, 1, 0)	(0, 0, 1, 0)	0
(0, 0, 1, 0)	(0, 0, 1, 0)	0
		total 5

a set in S'	a set in S closest to it	distance apart
(1, 0, 1, 1)	(1, 0, 1, 0)	1
(1, 0, 1, 0)	(1, 0, 1, 0)	0
(0, 0, 0, 1)	(1, 0, 0, 0)	2
(1, 0, 0, 0)	(1, 0, 0, 0)	0
(0, 0, 1, 0)	(0, 0, 1, 0)	0
		total 3

The distance between conceptualizations S and S' is 5 + 3 = 8.

Figure Captions

1. An object made from the Fischer-Technik assembly kit. It consists of 80 pieces and 104 physical connections.
2. An abstract graph of the object in Figure 1. Nodes represent pieces and links physical connections. Nodes are numbered 1 to 80 to indicate specific pieces in the object. Some of these pieces are noted in Figure 1.
3. The most typical conceptualization (division into sub- and sub-subassemblies) of the object in Figure 1. Derivation of this conceptualization is given in the text.
4. A minority conceptualization of the object in Figure 1. Among 47 trials, it is maximally distant from the typical conceptualization in Figure 3.
5. Names for the sub- and subsubassemblies in the typical conceptualization, as derived in Experiment IIa.
6. Names for the sub- and subsubassemblies in the minority conceptualization, as derived in Experiment IIa.
7. For each subject's conceptualization in Experiment IIb, we calculated its distance from the conceptualizations (perfect typical and perfect minority) shown in the videotapes. We projected each conceptualization down on the line from perfect minority to perfect typical, scaled from -100 to 100. This projection was called an x-score. The mean x-score for those viewing the typical videotape was 84.2. For those viewing the minority videotape, it was 51.2.
8. An example of clustering. It is explained in Appendix 2.

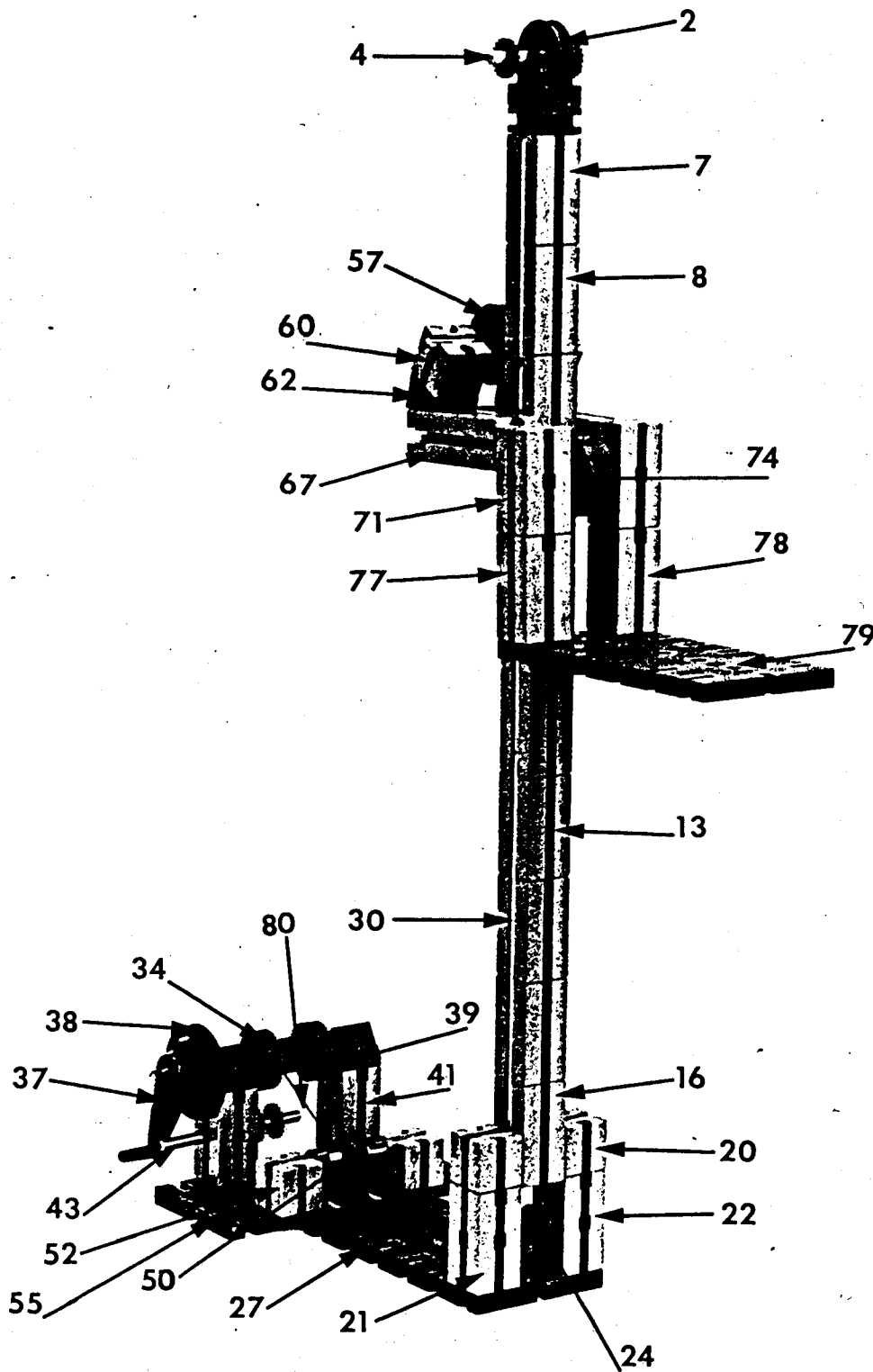


Figure 1

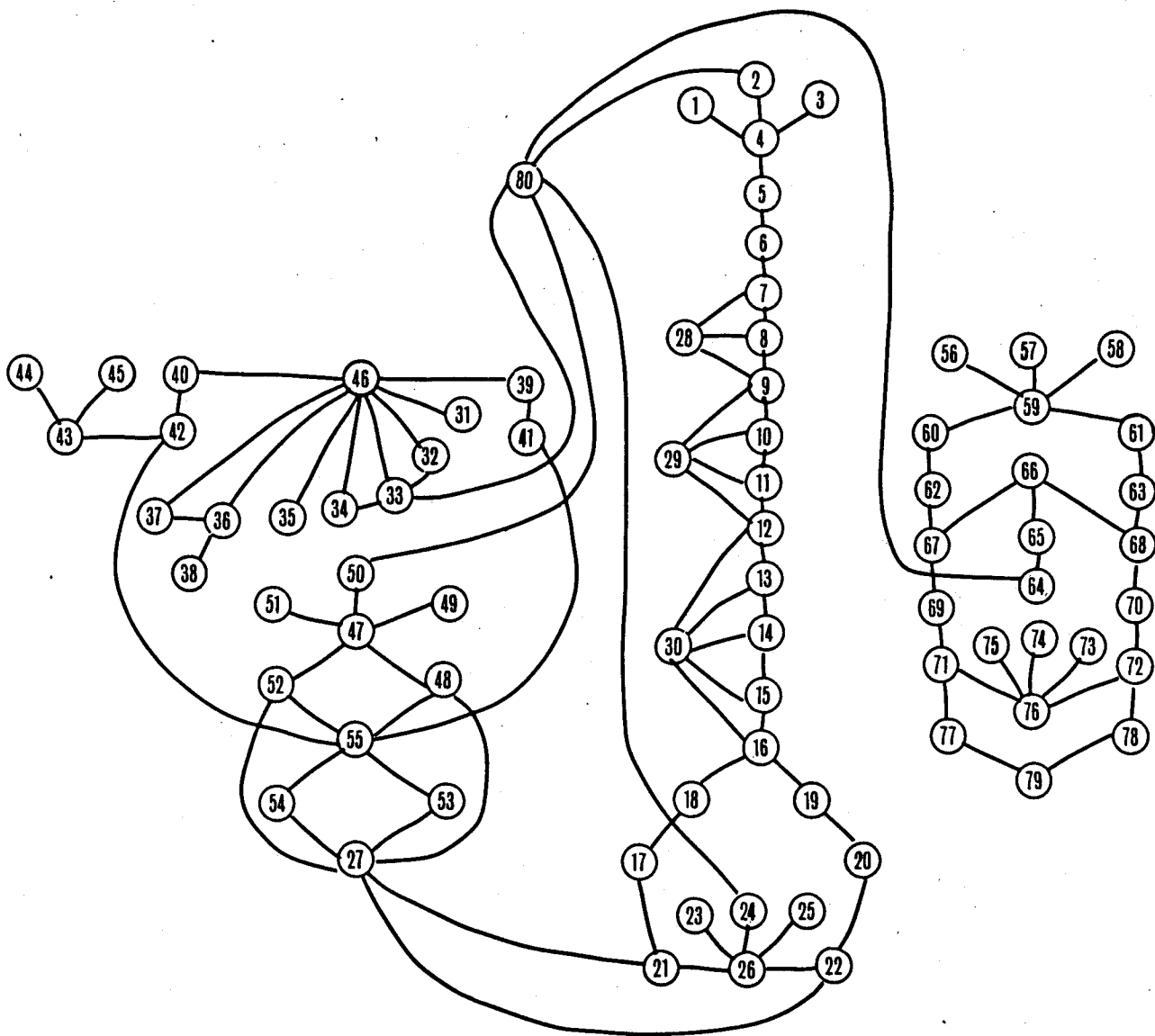


Figure 2

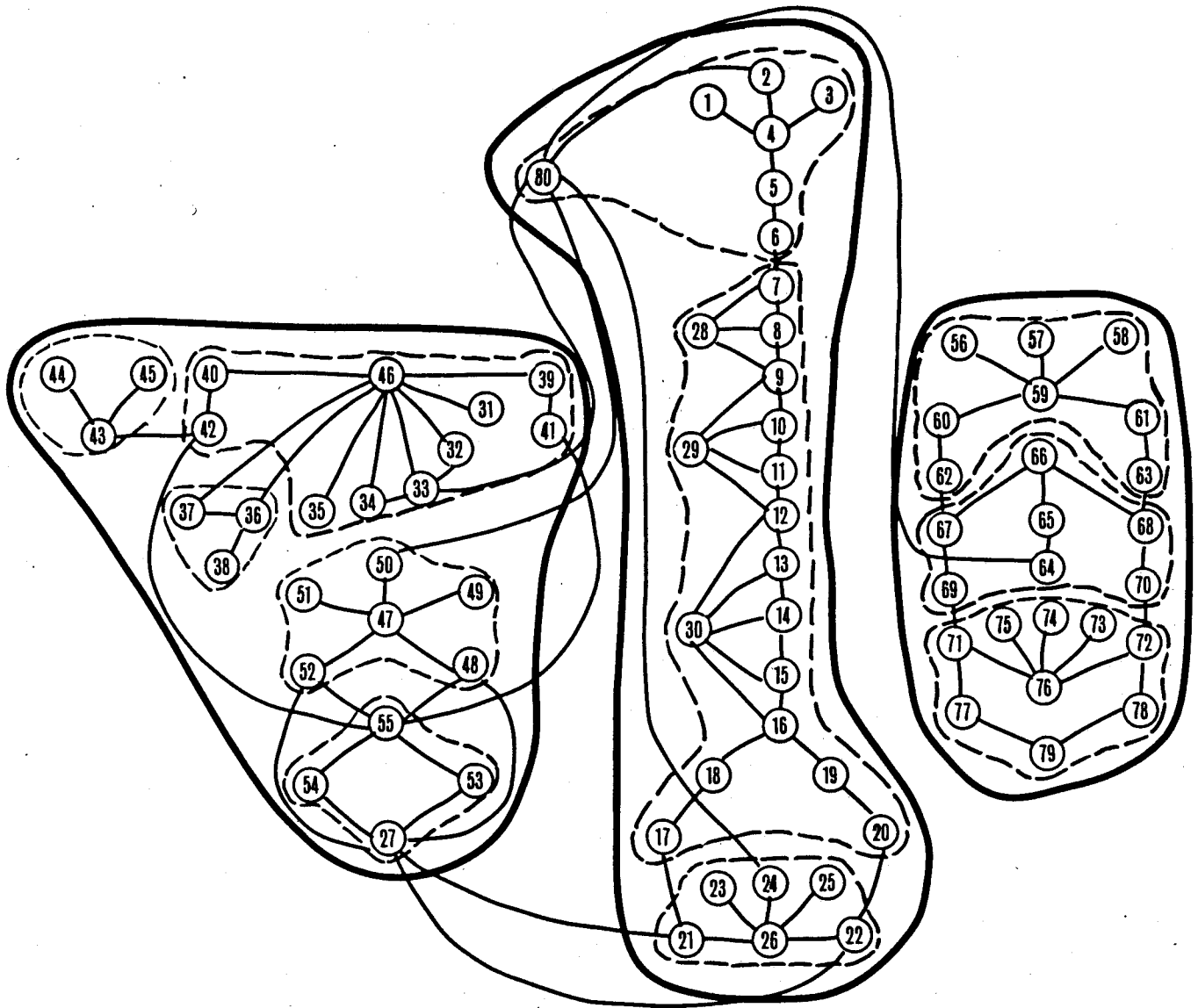


Figure 3

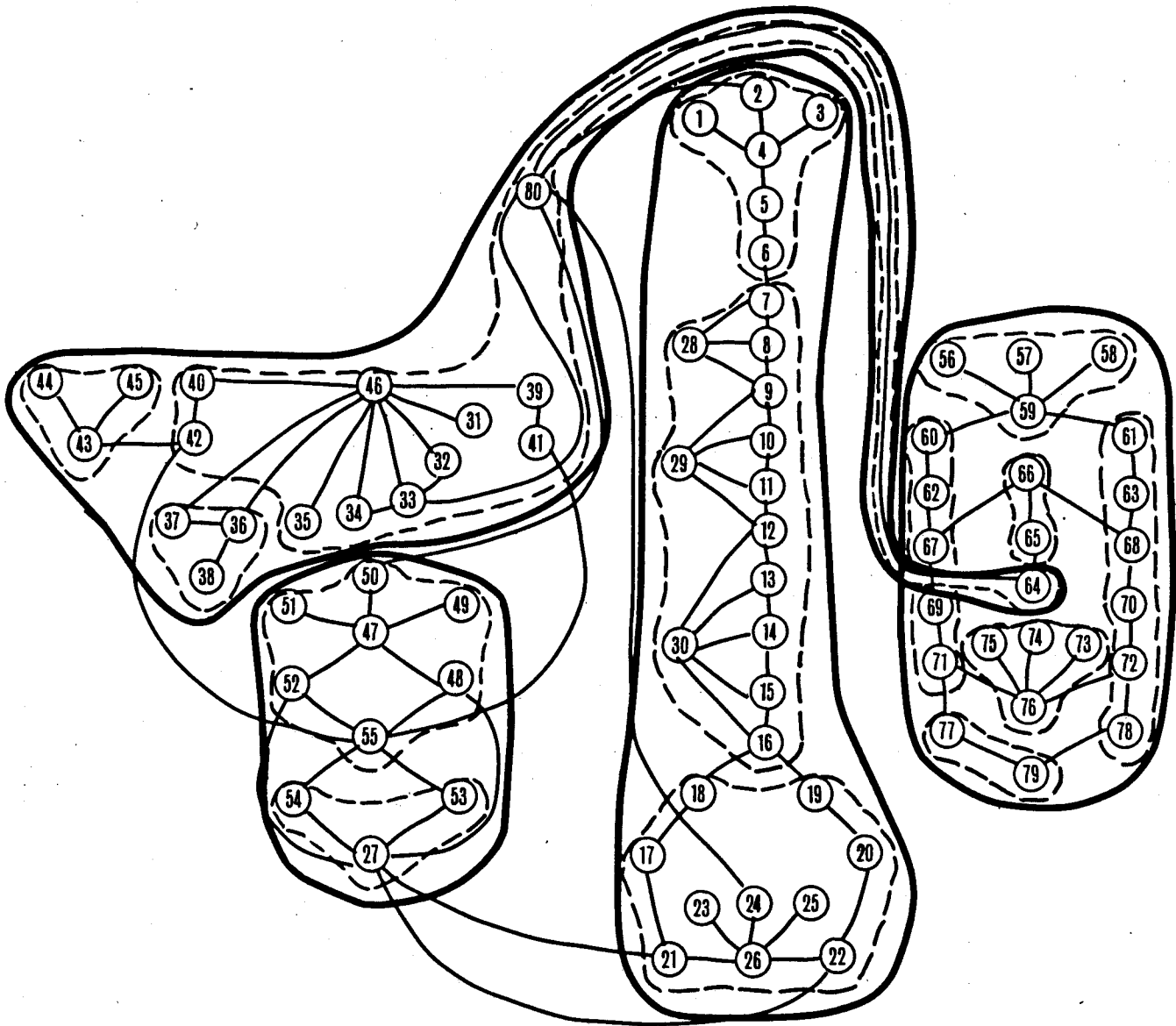


Figure 4

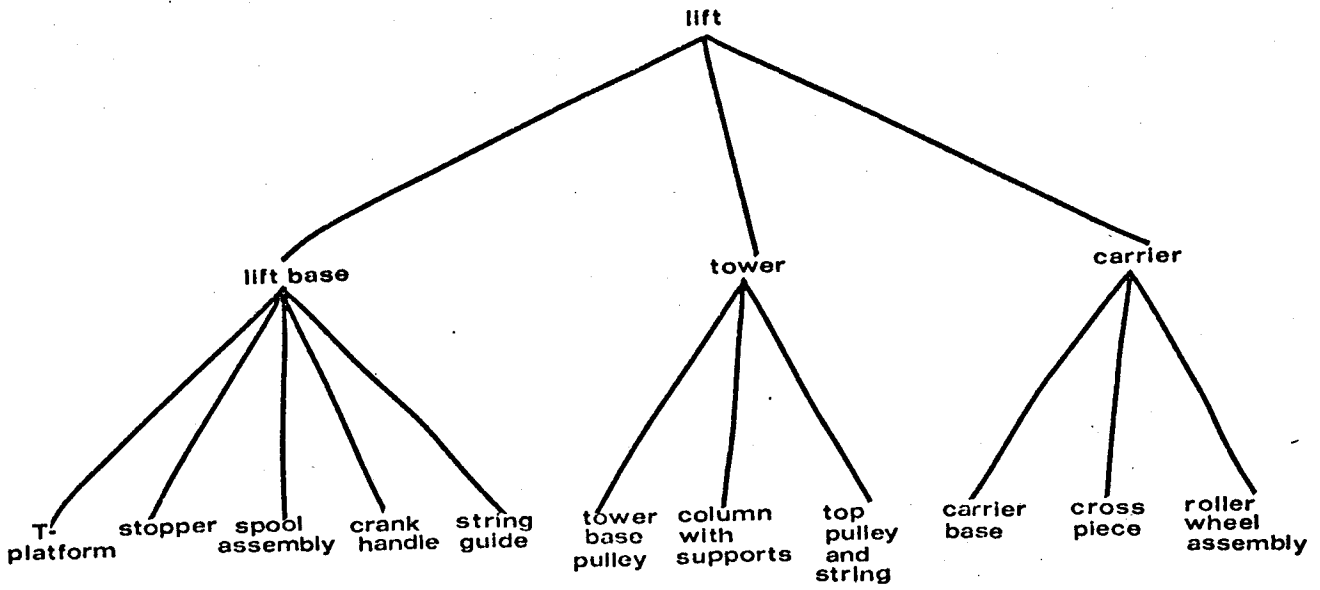


Figure 5

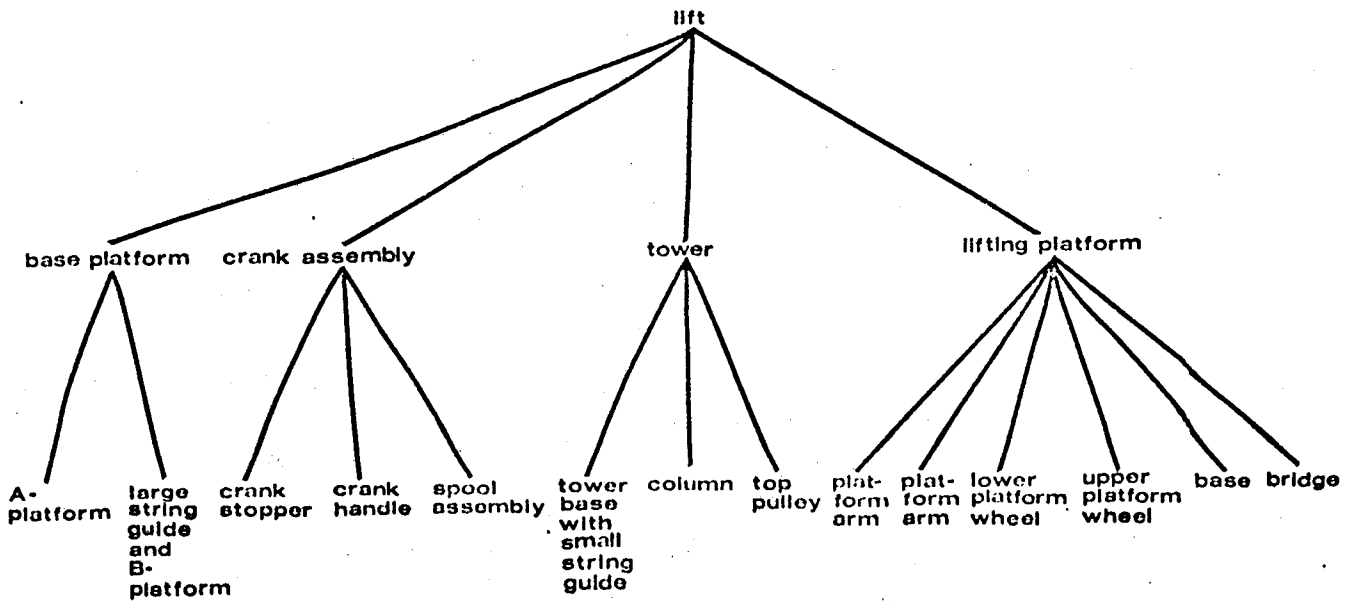


Figure 6

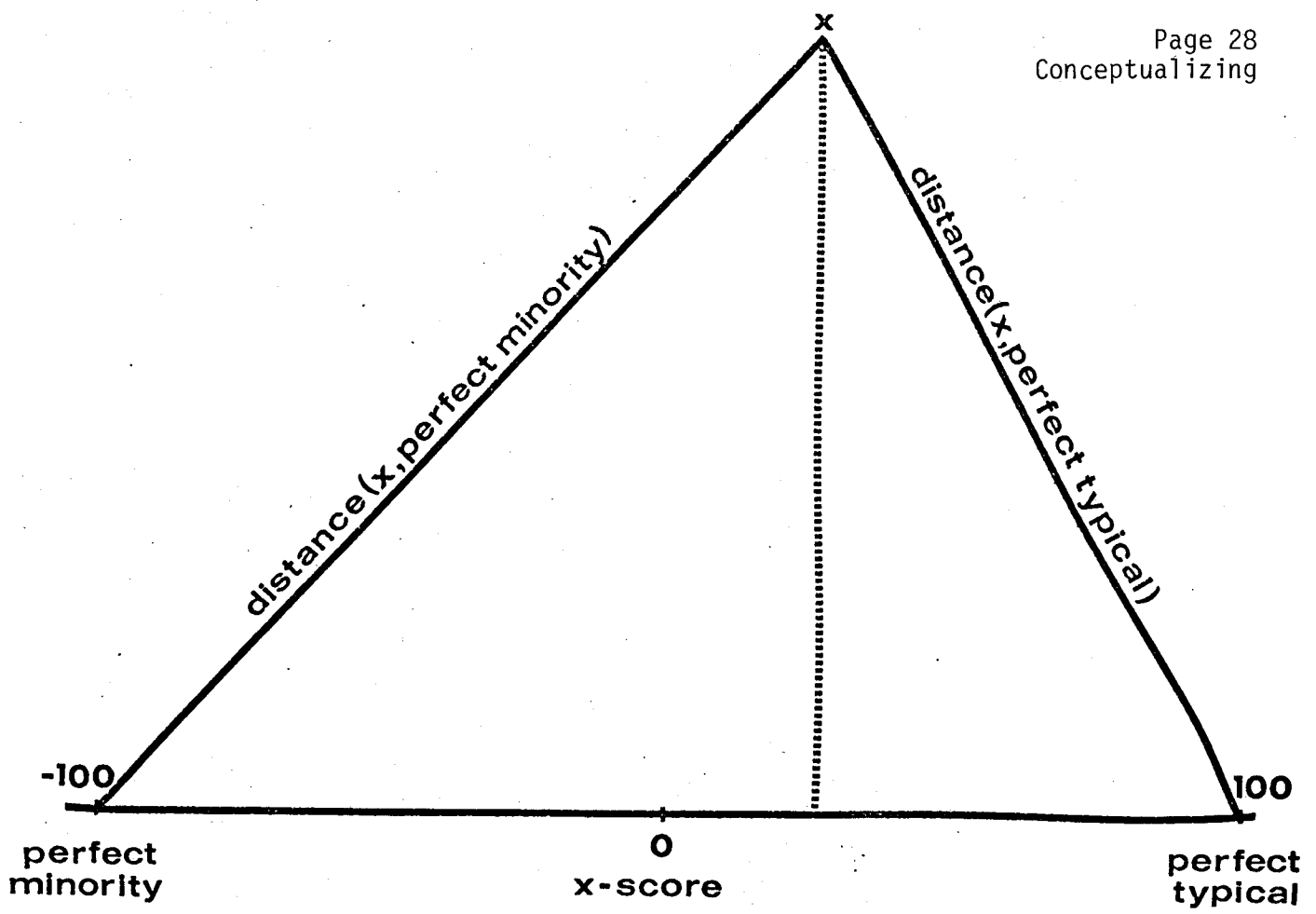
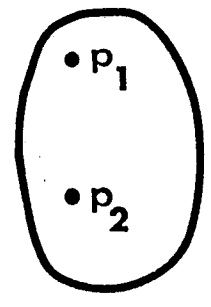


Figure 7

In conceptualization T_i :



In conceptualization T_j :

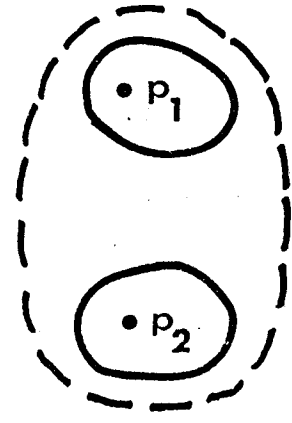


Figure 8

conceptualizations and the methods of measuring them a practical principle, and a way to implement it, are found. The principle: When a single set of procedural instructions is designed, it should present the conceptualization that the majority of people to be instructed by it bring to the situation naturally.