DOCUMENT RESUME

ED 072 974 SE 015 680

AUTHOR Johnson, Paul E.
TITLE Perceptual Structure of Scientific Knowledge. Final Report.
INSTITUTION Minnesota Univ., Minneapolis.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Div. of Elementary and Secondary Research.
BUREAU NO BR-0-0067
PUB DATE Sep 71
GRANT OEG-0-70-2650(508)
NOTE 134p.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS *Cognitive Development; Colleges; Educational Research; *Perception; *Physics; Research Methodology; *Semantics; Task Analysis; Transformations (Mathematics)
IDENTIFIERS Research Reports

ABSTRACT

The research contained in this report is directed at the problem of assessing structures of knowledge in science. Toward this end graduate as well as undergraduate students were asked to make judgments of the perceived similarity of expressions that mark or label concepts in the subject matter of physics. An apriori model of the structure of subject matter knowledge was used to generate task materials as well as interpret the judgment data. Evidence presented in the report lends suppport to the notion that mastery of a domain of knowledge can be indexed by procedures which might also reveal the state of a student's knowledge early in this educational experience. The adequacy of this approach to educational assessment (essentially through the medium of construct validity) should be tested by further research. Especially important, is research which reveals the role knowledge structure such as those identified in this report can play in problem solving behavior. (Author)

FINAL REPORT
Project No. 0-0067
Grant No. OEG-0-70-2650(508)

PERCEPTUAL STRUCTURE OF SCIENTIFIC KNOWLEDGE

Paul E. Johnson
University of Minnesota
Minneapolis, Minnesota

September 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

FINAL REPORT

Project No. 0-0067
Grant No. OEG-0-70-2650(508)

PERCEPTUAL STRUCTURE OF SCIENTIFIC KNOWLEDGE

Paul E. Johnson
University of Minnesota
Minneapolis, Minnesota

September 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Perceptual Structure of Scientific Knowledge

Table of Contents

ACKNOWLEDGEMENTS

As with most things, the research represented in this report reflects the efforts of many people. I am particularly grateful for the opportunity to work with the students who have found this project of interest. Among them are Tom Curran, Judy Garrard, Saied Jacob, Eugene Lenarz, Jed Lewis, Jim Reaves, and Larry Victor. I am especially indebted to David Cox who, from the very beginning, served more as a colleague than a student and to Steve Robbins who almost singlehandedly carried the project through to completion.

While it may not be readily apparent from reading the report, much of the data could not have been collected without the creation of a considerable amount of technical apparatus. The engineering skills of Floyd Rasch as well as his unfathomable patience, made our ideas possible. Even then, we should not have succeeded without Rod Rosse. His insights and ability to create computer data systems made possible the collection and analysis of large amounts of technical information.

Lastly, I have been fortunate to work with skilled secretaries who have managed my own activities as well as the activities of my staff. We are particularly grateful to JoAnne Glancy and Mary Larson whose efforts were never unappreciated.

Summary

The research contained in this report is directed at the problem of assessing structures of knowledge in science. Toward this end graduate as well as undergraduate students were asked to make judgments of the perceived similarity of expressions that mark or label concepts in the subject matter of physics. An apriori model of the structure of subject matter knowledge was used to generate task materials as well as interpret the judgment data.

Evidence presented in the report lends support to the notion that mastery of a domain of knowledge can be indexed by procedures which might also reveal the state of a student's knowledge early in his educational experience.

The adequacy of this approach to educational assessment (essentially through the medium of construct validity) should be tested by further research. Especially important, is research which reveals the role knowledge structure such as those identified in this report can play in problem solving behavior.

Chapter 1

Introduction

Much of our educational experience consists of regular attempts to assess intellectual capability and achievement. Very often, however, our assessment procedures lack firm psychological support. Evidence of this can be found in the current intelligence testing controversy and also in the difficulty facing those who must evaluate the outcome of instructional innovations. The research described in this report represents an effort to assess structures of knowledge underlying intellectual competence. More specifically, we wish to understand what a person knows when he has achieved sufficient status in our educational system so that we can infer he is modestly competent in some domain of knowledge.

An unfortunate feature of intellectual competence is its unavailability for examination. In problem solving, for example, we are typically able to examine only the outcomes of concentrated mental effort. In order to formulate our problem for experimental study we assume that knowledge can be conceptualized in terms of environmental variables rather than in terms of underlying psychological processes. This assumption allows us to interpret the complexity of behavior by means of relations in a stimulus field rather than in terms of interactions in a performance model of behavior.

We have also assumed that the search for powerful hypotheses about knowledge requires that we look to other disciplines for formulations that capture the complexities inherent in our intellectual experience. Models generated in the philosophy of science can help us understand the nature of scientific knowledge. Likewise, models from linguistics can help us understand behavior which utilizes language as a vehicle for storing and transmitting large amounts of technical information. The work outlined in the present report is based upon methodological knowledge from psychology; its propositional knowledge, however, has been drawn from fields of physics, philosophy of science, linguistics, and mathematics.

Whenever we use language as a vehicle for understanding behavior, we run the risk of finding out only what individuals know about language. That is to say, patterns of usage mask the underlying causes of behavior. This depends,

of course, on the tasks we employ. For example, writing sentences is different from verbal free association. But in general, we risk observing only those regularities which result from language rather than the knowledge which it is used to convey. We have a much better chance of finding out what an individual knows if we can utilize tasks such as psycho-physical judgment.

When an individual judges the similarities between two objects, we assume he capitalizes on complexities that exist in his mental or cognitive structure due to training. This view is based upon the further assumption that mental structure is developmentally the same for most people. What makes one person different from another is the individual training experiences that extended or complicate pre-existing structure. We suppose, for example, that the basic cognitive structure shared by all psychologists is a distortion of some more fundamental structure about human behavior which is shared by all members of a given cultural-linguistic community. By this same reasoning we suppose that the cognitive structures shared by all operant conditioners is a distortion of the structure shared by all psychologists. This assumption about the nature of cognitive structure leads to a view which allows us to talk about how individuals differ in terms of what makes them the same; it is a guiding principle of the research described in this report.

The goal of the research reported here was to arrive at an understanding of specific structure of knowledge in science. One outcome of this type of work is the accumulation of normative information which can be used to diagnose ineffective or deficient performance as well as assess mastery. A second, more general, outcome is the development of knowledge that will illuminate our understanding of the psychological foundations of educational assessment. A third and more obscure outcome is the achievement of insight into the psychological foundations of science itself.

Because the report represents a somewhat unusual combination of physics, mathematics, and psychology, a word is in order regarding the manner in which it has been organized.

Chapter 2 contains the theoretical structure underlying our work. We begin with an overview of assumptions regarding the nature of scientific knowledge. These assumptions form the basis for the model used to describe knowledge for purposes of psychological experimentation. The model can be interpreted as a

grammar for the language of physics since it produces all and only those expressions in the subject matter, and since it is capable of assigning a structural description to each expression based upon its history of generation. Chapter 2 concludes with a description of the model as well as the mathematical scheme from which it derives.

Chapter 3 begins with a discussion of the nature of psychological experimentation undertaken to examine the role of perceptual structure in the formation and utilization of scientific knowledge. We then consider two experiments which serve as prototypes for the rest of the report. The methodology, as well as the findings from these experiments are developed in detail. Chapters 4-7 present the bulk of our findings. Chapter 8 is an epilogue which attempts to place the findings of the project in the perspective of increasing efforts to make the science of psychology relevant to educational practice.

Chapter 2

The Theory

We shall suppose that scientific (physical) theory consists of three parts (Carnap, 1939; Nagel, 1961). There are, first, the rules or equations of the theory such as Newton's equations of motion in mechanics (e.g., $F = ma$). These equations do not contain the means for determining whether they are in agreement with physical facts or whether they are logically consistent. That is to say, the equations of mechanics do provide any evidence of their own logical consistency or physical truth. The source of logical consistency for these equations is a second set of rules which, in the case of physics, consists of algebra and geometry. However, equations of motion plus the laws of algebra and geometry are not sufficient to establish the equations as physical laws. A third system of rules, called semantic rules, is required. These rules serve to connect equations in the theory with words in everyday language. By adding semantic rules to the equations of motion, equations such as $F = ma$ become physical laws testable by experimentation (Frank, 1946).

Theories in science confirm meaning and significance upon experience by providing the language necessary to talk about that experience (Feynman, 1965). Semantic rules are necessary because the formal structure of a theory says nothing whatever about the world. A formal theory is a decision about the use of terms. To propose a definition is not necessarily to advance knowledge, but merely to establish rules of usage. The semantic rules in science state the conditions which happen to be satisfied by anything to which the terms in the theory might apply. By changing semantic rules we change what there is to talk about.

The propositions of a theory in science are often related and described by a model (e.g., the Bohr model of the atom). Such a model determines the domain of application of the formalism by suggesting the appropriate semantic rules Models are used in this sense to illustrate the structure of concepts in a theory. In addition, perceptions, which determine how activities (e.g., problem-solving) are carried out typically embody the current models for interpreting the experience to which the theory addresses itself (Kuhn, 1962). Very often such models are spatial or mechanical in nature and the perceptions which guide activity in the science are similar to those found in common sense or prescientific thinking (Jammer, 1957).

The equations of physical theory may be written at various levels of abstraction. At the highest level is something like F = ma which is not so much an equation as a schematic form whose precise symbolic representation varies from one application to the next. In the case of free fall, for example, F = ma becomes $-mg = m\frac{d^2x}{dt^2}$. In the case of the simple harmonic oscilator the equation F = ma is written as $-kx = m\frac{d^2x}{dt^2}$. Abstract equations such as F = ma are the general equations from which expressions embodied in particular problems such as a falling body and the harmonic oscillator can be derived. It is these latter expressions which allow a theory to be related to phenomena and which lead to prediction.

Semantic content is embodied in prototype problems which are learned by all members of a scientific community. Such problems are usually found at the end of chapters in scientific texts so that students can learn how higher order abstract equations within physical theory are given particular form to describe certain classes of natural phenomena.

Among the things which are learned by individuals when they have become proficient in science are similarity relationships. These relationships (a) relate basic concepts to one another and (b) form the basis for viewing new phenomena as instances of prototype problems. We suppose the network of interrelations among concepts guides the use of words and symbols in the language of science while in the case of problem solving the scientist discovers how to see a particular problem as similar to a problem he has already encountered. Once a specific problem is seen as an instance of some prototype (e.g., the simple harmonic oscillator), the mathematical means for solving the prototype can be used to solve the specific problem Prototype problems are the scientific community's standard examples, and it is these problems which form the basis for scientific activity and communication.

Most research on knowledge has focused upon the skilled behaviors which are more or less agreed upon as indicative of knowledge of a concept of group of concepts. These behaviors are arrived at by asking specialists in the subject matter "what should people who know a concept be able to do." This approach has its origins in training research and research on operant behavior (programmed

instruction) and leads, under the best of circumstances, to a domain of criterion behaviors. This domain defines what it means behaviorally to know the subject matter. From such research we know a good deal about the conditions under which specific problem solving skills are learned and maintained (cf., Gagne, 1962, 1967; Hively, 1968). Much less is known, however, about the acquisition of rules which determine the applicability of these skills and knowledges, and yet it is just these rules which are at the heart of the communication and understanding of scientific knowledge (Bohm, 1965; Deese, 1969).

It is an assumption of this research (and in fact a departure from tradition) that an adequate psychological understanding of scientific knowledge cannot come solely from answers to the question, "What should an individual be able to do", but must also include experimentation based upon independent variables arrived at by a logical analysis of the concepts involved. In this report we shall describe scientific knowledge in terms of a deductive system consisting of physical equations, a mathematical system which tests the consistency of these equations, plus a set of semantic rules which relates symbols in the equations to everyday language. At the heart of this system is the idea of logical or mathematical generation.

A recursive definition is a common mathematical tool which allows one to state simple rules or procedures which can be applied repeatedly, thus building a complex structure by means of many small steps. Given an initial set of primitives and some artibrarily selected stopping point, the rules of generation determine what will be produced.

We use the idea of a recursive definition to generate expressions in the language of analytical mechanics. This content was chosen for several reasons: (1) because its elements are reasonably stable and well-defined, (2) because it is characterized by well-defined prototype problems which contain examples of the phenomena it describes, and (3) because it uses the language of mathematics as a vehicle for transmitting and storing information.

One means of conceptualizing our procedure is to suppose that our goal is to produce items which test for subject matter knowledge. Figure 1 is a device for visualizing the framework within which our model is constructed. Figure 1 begins at the top of the page with what we call cognitive categories. These categories are fundamental or prerequisite for behavior. Some examples of such categories would be

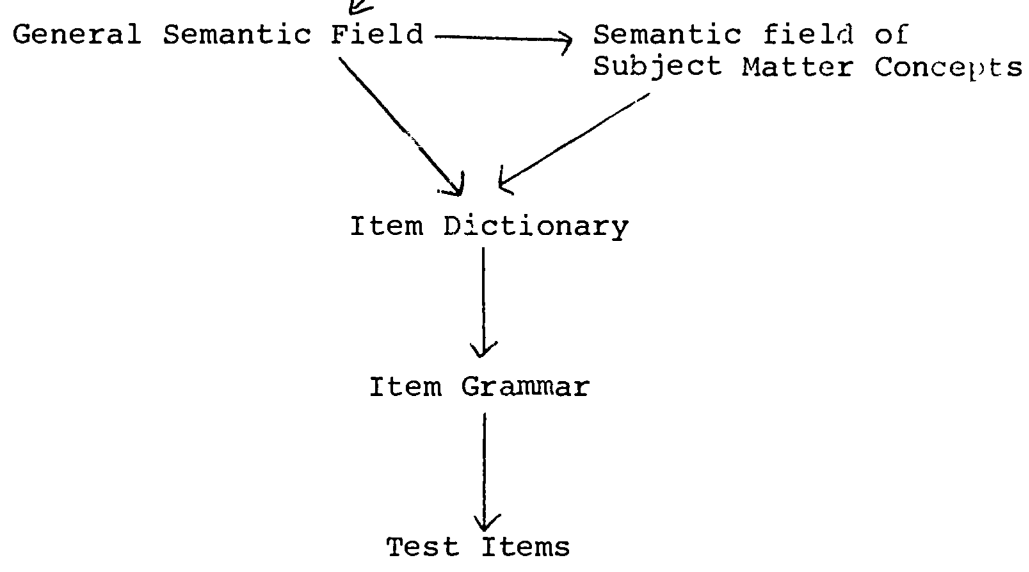


Fig. 1. Model for the Generation of Test Items

to a categorical dictionary which contains elements such as mass, distance, time and force.

There are two paths in the diagram in Fig. 1. In one direction we construct a set of rules which, when applied to points in the semantic field, results directly in an item dictionary. In the other direction we apply rules to points in the semantic field to determine which are subject matter concepts. This latter direction involves asking whether expressions represented by points in the semantic field are true of some physical system. The procedures for making this test are described shortly. Suffice it to say for now that the result of this testing is still a set of points in a model space. Only now the model space has a more complex dimensionality.

In either case, we construct the item dictionary by applying rules to points in a semantic field. These rules are simple propositions (if-then statements such as, if x_2 is "a" and x_3 is "b" what is x_1). The item dictionary is analogous to the categorical dictionary, except that in this case its elements are abstract propositions.

The item dictionary can be conceptualized as a configuration space consisting of nt dimensions where n is the number of points sampled from the semantic field and t is the dimensionality of the model space for the semantic field. Points in this space represent all the points in a given sample from the semantic field. Successive samples from the semantic field become individual points in configuration space.

We can conceptualize the successive sampling of points in order to generate elements of the item dictionary by a graph such as that presented in Figure 2. The "trajectory" of the graph embodies the particular

Table 1

Experimental Stimuli for Each Set

Vertex in Figure 3	Stimulus Set αβγ	mxt	gm-cm	color[a]
1	$d\beta/d\gamma$	dx/dt	cm/sec	SR∆
2	$\alpha d\beta/d\gamma$	mdx/dt	gm-cm/sec	LR∆
3	$\alpha d^2\beta/d\gamma^2$	md^2x/dt^2	gm-cm/sec^2	LR□
4	$d^2\beta/d\gamma^2$	d^2x/dt^2	cm/sec^2	SR□
5	$\int(d\beta/d\gamma)d\beta$	$\int(dx/dt)dx$	cm^2/sec	SY∆
6	$\int(\alpha d\beta/d\gamma)d\beta$	$\int(mdx/dt)dx$	gm-cm^2/sec	LY∆
7	$\int(\alpha d^2\beta/d\gamma^2)d\beta$	$\int(md^2x/dt^2)dx$	gm-cm^{2}sec^2	LY□
8	$\int(d^2\beta/d\gamma^2)d\beta$	$\int(d^2x/dt^2)dx$	cm^2/sec^2	SY□

[a]Stimuli in the color set were colored geometric forms; they are represented here according to the following scheme:

S = small R = red ∆ = triangle

To control for order effects across stimulus sets, each S received all four sets in a different random order, although, for any given set, the doubles task always followed the singles task. Two random orders of stimulus presentation were constructed for both the singles and doubles tasks for each stimulus set; stimuli which appeared on the left of the screen in order 1 appeared on the right of the screen in order 2. Ss were alternately assigned to one of the two random orders for each stimulus set.

Ss were 17 male graduate physics students from the University of Minnesota. They were tested individually and paid six dollars each for participating in the two hour experiment.

Results and Discussion

The first question considered in the examination of data was how well Ss' judgments within each stimulus set conformed to the equality relation implied in Fig. 3. This relation specifies that the four stimulus pairs within a category such as p edges, q edges, pq diagonals, etc. should be judged identical in dissimilarity.

Means and standard deviations were computed for Ss' judgments of each stimulus pair in each set (see Table 2). Inspection of the standard deviations indicates a considerable range of variability among Ss within each stimulus set. Inspection of the means for all stimulus sets indicates that averaged judgments within any one category were reasonably alike (although not identical),

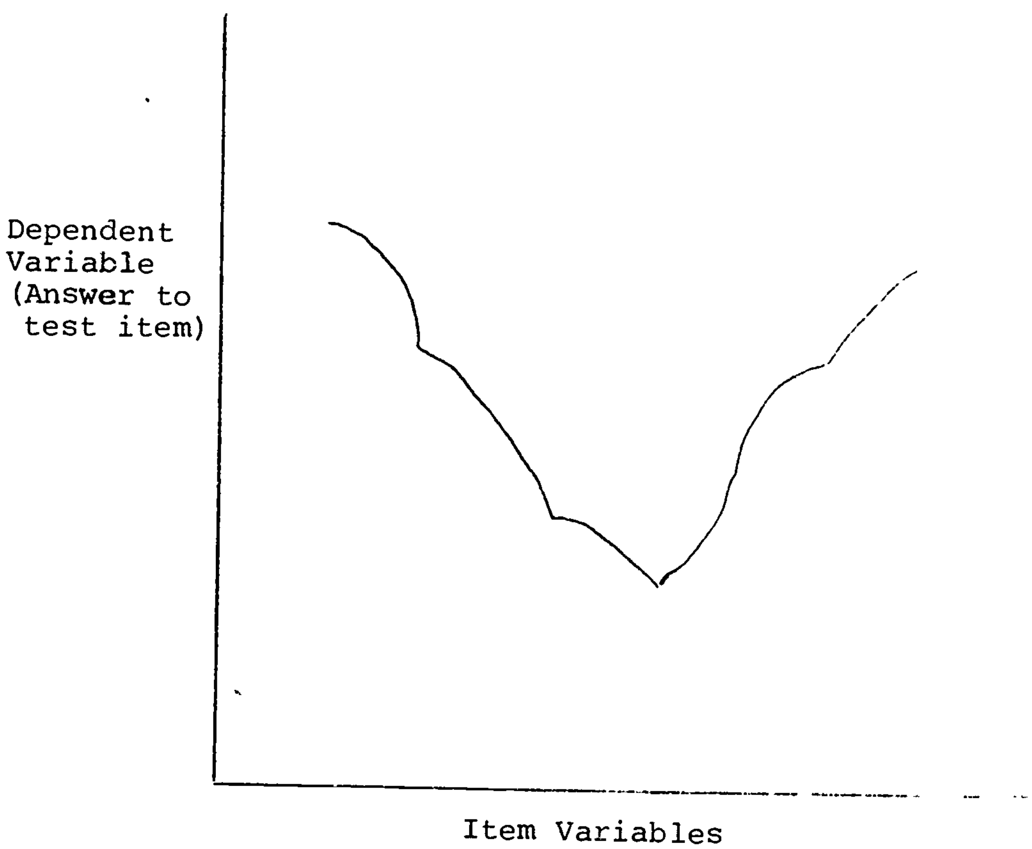


Fig. 2. Graph representing Questions Generated for a Particular Set of Points in the Semantic Field.

set of questions that we have chosen to ask about concepts represented in the semantic field.

According to Fig. 2, there is a certain minimum number of categories which must be known in order for the value of the dependent variable (answer to the test item) to be determined. There are local minima where we are given what is required but several additional steps are needed in order to arrive at the value of the dependent variable. The absolute minimum represented by the lowest point in the graph is a relationship which defines a particular concept. For example, if we are generating test items about the concept of force, the minimum might be represented by force = mass x acceleration.

It is also possible to interpret the minimum number of variables in any test probe by the shortest distance in model space between concepts. Distances between points in model space can be defined in two ways: either logically if we have no a priori psychological information, or psychologically by particular psychophysical functions, if we have experimentation to determine the nature of the psychological distance between points.

The question of difficulty with regard to test items can be thought of as a function of the distance between points in the semantic field. The least difficult test items, for example, should be those which are represented by the absolute minimum in Fig. 2. As we move away from this minimum, test items become more and more difficult. Put another way, the items which minimize $s = s\ (d_1, d_2, \ldots d_n)$ will be easiest to solve.

To apply our scheme of analysis to concepts in mechanics we begin with the mathematical generator. The generator operates upon primitives (e.g., m, x, t and F) and converts them into mathematical expressions. Some of these expressions are found in analytical mechanics and many are not. All such expressions can be represented as points in a semantic field.

Fig. 3 represents a semantic field used extensively throughout the project research. Fig. 3 portrays alphanumeric expressions in physics as points in a Cartesian (Euclidean) 3-space. The axes of the space represent calculus operations with respect to m, x, and t (mass, distance and time, respectively). Distances between points in the space are assumed to correspond with the nominal (or logical) similarity among expressions viz., the shorter the distance, the greater the similarity among

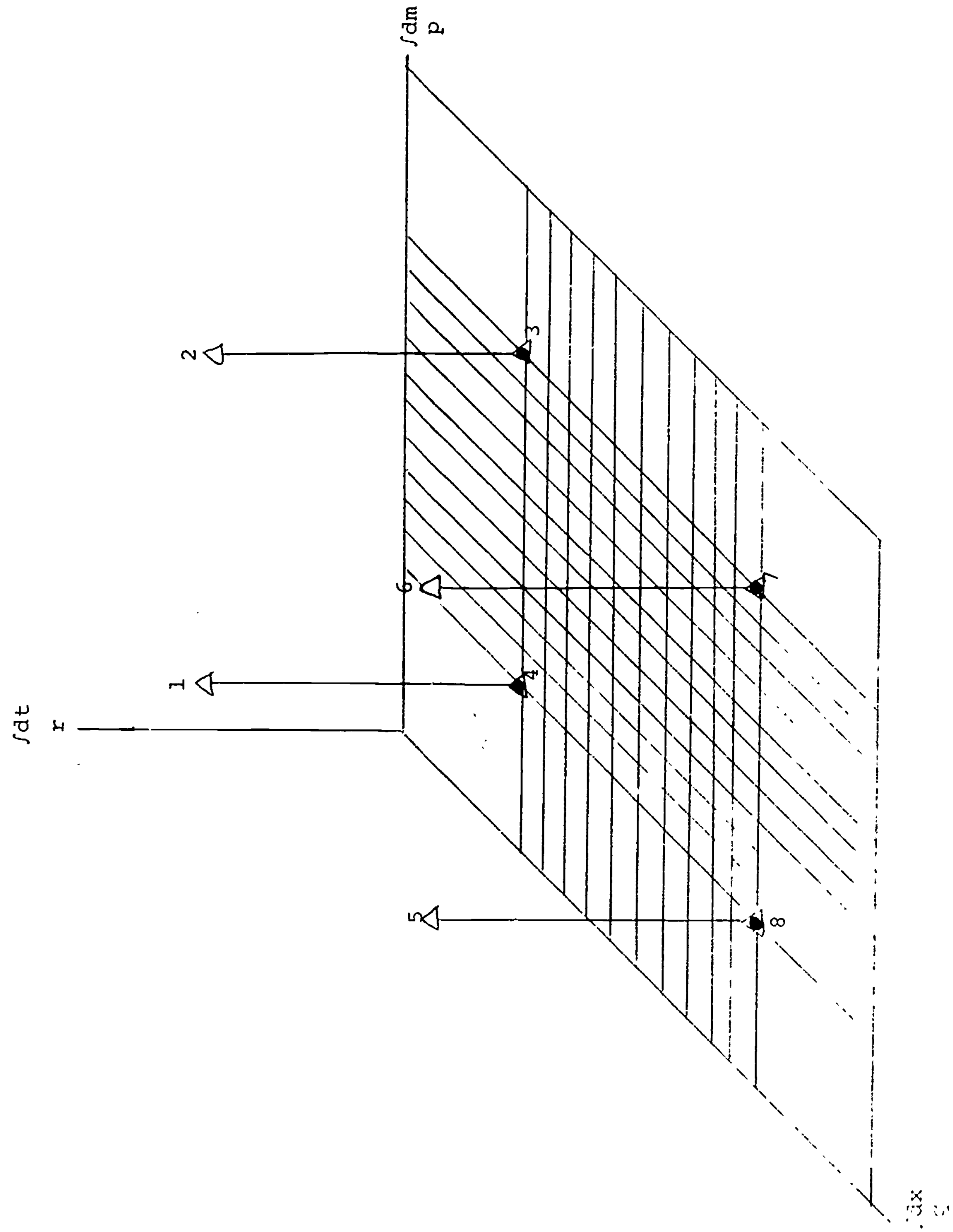


Fig. 3. Model for describing a structure of knowledge in physics.

expressions represented by those points.

The model is used to generate expressions by starting at some point in the space with an operand and applying the calculus operations indicated by the axes. Motion in a positive direction on the p axis indicates integration with respect to mass ($\int dm$). Motion in a positive direction on the q axis denotes integration with respect to distance ($\int dx$). And motion in a positive direction on the r axis indicates integration with respect to time ($\int dt$). For each axis, motion in the negative direction indicates differentiation (e.g., motion in the negative direction onthe r axis indicates differentiation with respect to time, d/dt).

If, for example, one starts at point (1) with the operand dx/dt, and solves all integrations with respect to mass as indefinite integrals (i.e., $\int\frac{dx}{dt}dm = m\frac{dx}{dt}$), then the model can be used to generate the following eight expressions:

(1) $\frac{dx}{dt}$, (2) $m\frac{dx}{dt}$, (3) $m\frac{d^2x}{dt^2}$, (4) $\frac{d^2x}{dt^2}$, (5) $\int(\frac{dx}{dt})dx$, (6) $\int(m\frac{dx}{dt})dx$, (7) $\int m\frac{d^2x}{dt^2}dx$, (8) $\int(\frac{d^2x}{dt^2})dx$. Five of the expressions produced in this fashion are readily labeled by concept words in physics and three are not. The labeled expressions are: (1) velocity, (2) momentum, (3) force, (4) acceleration, and (7) work (energy). The fact that the model allows for the generation of unlabeled points as well as labeled points can be interpreted to mean that knowledge of its structure should allow one to interpret any expression that is written according to its rules (much like one interprets new sentences in a language).

The model defines content markers (concepts) by the postulation of primitives, and so leads to a deductive system for describing the content domain of classical physics. While not without its difficulties, such an approach is useful to the psychologist in that it provides a tidy means of organizing content in a manageable framework for purposes of study and experimentation. There is also a view which suggests that deductive systems may be usefully applied to content domains as a way of generating new content in an orderly fashion (Tisza, 1962).

For purposes of describing the relationships among expressions generated by the model, we shall assume that all calculus operations along a given axis in Fig. 3 are spatially equivalent for any operand (i.e., the geometric representation obtained by connecting pairs of points parallel to an axis is a regular polyhedron). Moreover, since we have no a priori information about the relative importance of the dimensions in the model, we shall assume that integrations along different axes are spatially equivalent, so that the figure representing relationships among the eight expressions may be drawn as a cube. It should be understood, however, that the interrelationships embodied in Fig. 3 represent an idealized structure rather than a psychological structure of interrelations such as might be found in the performance of Ss on some task.

Because the language of science is often highly mathematical, it is reasonable to assume that scientists, and students of science, code and recognize concepts largely on the basis of their mathematical form. It is also likely that concepts have semantic as well as mathematical (syntactic) features. While recognizing that the issue of meaning in science is complex, and that many terms in the language of science have only partially determined meanings (Bar Hillel, 1969; Bloomfield, 1939; Carnap, 1966), we shall assume that the syntactic features of expressions in physics are embodied in mathematical operators and that semantic features for a given syntax are portrayed by specific combinations of operands.

To determine which expressions belong in mechanics, the expressions must be combined into equations. This last step is necessary because any expression is possible in mechanics; only equations can be accepted or rejected. The construction of equations is accomplished by pairing expressions with one another using an equal sign. Those expressions which have mathematical and unit consistency are then tested to see if they are instances of one of Newton's general laws. That is, a true pair of expressions (an equation) is defined as $x = f(t,\ldots)$ $F = g(t, x\ldots)$ where $F = \frac{md^2x}{dt^2}$.

All pairs of expressions which are related in this sense are semantically possible statements in the subject matter (i.e., they represent concepts). To complete our description of mechanics the pairs of expressions must be sorted according to prototype problems and ultimately the problem instances to which they apply.

Generation of Test Items

Test items can be constructed by following either one of the two paths indicated in Fig. 1. In taking the first path, we label the points in Fig. 3 by reference to basic definitions in the subject matter. Thus, as stated earlier, velocity = 1, momentum = 2, force = 3, acceleration = 4, work (energy) = 7. If we now adopt a rule of the form, given x_1 and x_2, what is x_3? We can write elements of an item dictionary by systematically permuting for x_1, x_2, and x_3 each of the above five concepts. We arrive at a configuration space for the item dictionary by keeping the dependent variable (x_3 in the above case) constant while permuting the other concepts for x_2 and x_1. We then obtain a new configuration space (new item dictionary) each time we change the dependent variable.

The final step is to choose a language in which to express the elements of the item dictionary. The grammar of this language is the vehicle by which elements are converted to test items. An example for the above concepts would be: Suppose mass is 10 and acceleration is 3, what is force? The procedure for this path is straight-forward and quite mechanical. While the resulting test items have little apparent "physical content" they do represent one aspect of the subject matter, namely, the mathematical interrelationship among the concepts. To generate items with other content we must take the second path in Fig. 1. In order to follow the second path we need a semantic field for the subject matter.

Consider the physical system consisting of

1) A single point mass which is free to move along a straight line in space;
2) A constant force parallel to the straight line.

The mathematical equivalent of the above is: $\bar{x} = [x(t), 0, 0]$ $\bar{F} \cdot Y = \bar{F} \cdot Z = 0$ and $\frac{\partial \bar{F}}{\partial t} = 0 = \frac{\partial \bar{F}}{\partial x}$ (where $\bar{F}$ denotes force as a vector quantity and $\frac{\partial \bar{F}}{\partial t}$ denotes the partial derivative of force with respect to time). Since force and mass are constants for this system, they can be prescribed as real numbers, (i.e., they are variables of the item dictionary). Once force and mass are chosen, the form of x is determined (i.e., force must equal $\frac{d^2x}{dt^2}$).

The specification of $\bar{F}$ and m also means that any equation from the generator which contains either term is totally or partially specified. For instance, we could choose

$$\begin{aligned}
&\text{acceleration} && = a(t) = F/m \\
&\text{velocity} && = v(t) = v_o + F/mt \\
&\text{distance (position)} && = x(t) = x_o + v_o t + 1/2\ F/mt^2 \\
&\text{Work} && \begin{cases} = U(x) = Fx_o - F_x \\ = U(t) = -F(v_o t + 1/2\ F/mt^2) \\ = T(t) = 1/2\ m(v_o + F/mt)^2 \end{cases} \\
&\text{momentum} && = p(t) = mv_o + Ft
\end{aligned}$$

where the subscript o refers to an initial value. This initial value is a property of the mathematics of the generator (i.e., it is a constant of integration) and may not be applicable in a given situation. If it is zero, the terms containing it vanish.

Applying the same item rule as before we obtain: If a point particle of mass 3 is moving with a velocity of 5 in space and is subjected to a force of 10 in the direction of its motion, what is the resulting acceleration of the particle? The only thing missing from such an item to make it acceptable to the subject matter specialist is units of measurement. These can be added in a straightforward manner by selecting the unit for the dependent variable and requiring that units be balanced on both sides of the statement or equation.

Summary

The scheme proposed in this chapter has both psychological and educational significance. Psychologically, it can be used to generate sets of nominally related stimuli for purposes of experimentation. Furthermore, the semantic field can be thought of as a model for subject matter competence. The explication of the precise psychological functions which define distance between points in this field may shed light upon basic processes in perception and cognition.

Educationally, the scheme allows test items to be generated from a few simple assumptions regarding the content and form of these items. The scheme also permits a definition of item difficulty and the isolation of potential sources of confusion among items, all of which can be tested empirically in a teaching-learning context.

References

Bar-Hillel, Y. Universal semantics and philosophy of language: Quandries and prospects. In Puhvel (Ed.), Substance and function of language. California: 1969.

Bloomfield, L. Linguistic aspects of science. International Encyclopedia of Unified Science, 1939, 1, 1-59.

Bohm, David. Physics and perception. The Special Theory of Relativity by David Bohm. W. A. Benjamin, New York, 1965, 185-230.

Carnap, R. Foundations of logic and mathematics. International Encyclopedia of Unified Science. 1939, 1, No. 3.

Carnap, R. Philosophical Foundations of Physics. New York, 1966.

Deese, J. Behavior and fact. American Psychologist, 1969, 24, 515-522.

Feynman, Richard. The Character of Physical Law. The MIT Press, 1965.

Frank, P. Foundations of physics. International Encyclopedia of Unified Science. 1946, 1, No. 7.

Gagne, Robert. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagne, Robert. Curriculum research and the promotion of learning. In R. Typer, R. Gagne & M. Scriven, Perspectives of Curriculum Evaluation. Chicago: Rand McNally & Co., 1967.

Hively, W., II, Patterson, H., & Page, S. A universe-defined system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-289.

Jammer, M. Concepts of Force. Cambridge: Harvard University Press, 1957.

Nagel, E. The Structure of Science. Harcourt, Brace & World Inc. New York, 1961.

Tisza, L. The logical structure of physics. Synthese, 1962, 14, 110-131.

Chapter 3

Demonstration of Perceptual Structure

To infer the structure underlying scientific expressions, we must have two or more measuring procedures for a given set of stimuli so we can assume that subjects use the same label for the same stimulus within a variety of contexts (Garner, Hake & Eriksen, 1956).

The major procedure used to assess relations among stimulus materials in the present research was rating scale judgments (e.g., subjects judged the similarity of two stimuli on an eleven point scale -- 1 meaning highly similar and 11 meaning highly dissimilar). We assume that an individual's response on a scale of numbers to pairs of stimuli index similarity relationships between these stimuli. We further suppose that subjects have little or no experience in using numbers on a scale to describe similarity among expressions and that an individual's responses are made only on the basis of perceived similarity, not on the basis of response characteristics such as an artifact of assigning numbers.

One method of interpreting similarity data (i.e., responses on a scale of similarity as defined above) to recover metric properties of structure is to assume that these responses can be interpreted as psychological distance in a psychological space. Stimuli are represented as points in a space having values on each of the dimensions which define the space. The distances between stimuli are computed using a distance function, and these distances can be used to predict responses in tasks which differ markedly from the original tasks used to construct the space.

We assume that individuals in a scientific community communicate more readily with one another to the extent they have similar or compatible psychological structures. That is to say, those individuals who are alike in their judgments regarding the similarity of stimuli should understand each other's communications about those stimuli more effectively than individuals who differ from one another in this respect (Fodor, 1968).

The tasks we have chosen in this project to test the hypothesis that similarity responses index structures in science, are communication tasks. In these tasks

subjects must (a) group stimuli on the basis of criteria such as physical "appropriateness" and (b) pick out the "inappropriate" stimulus from a group constructed so that all but one of its members are close to one another in similarity.

Background of Research

The research which led up to the project was done using associative and judged similarity as dependent variables for studying the structure of words in high school physics. These data were collected over a five year period and were originally part of an analysis of scientific communication on the relatively primitive level of introductory Newtonian mechanics. From this research we know that there are statistical variables in the informal language of scientific communication. These variables are functionally related to word association contingencies, response distribution characteristics, and similarity judgments (Johnson, 1964, 1965, 1967, 1969a, 1969b).

More recent research based upon a sample of subjects taken from a graduate school population in physics at the University of Minnesota has also demonstrated that associative similarity and judgments of perceived similarity among words in analytical mechanics reflect a structure for the concepts in the subject matter.

In a study by Johnson, Cox and Curran (1970) graduate students in physics served as Ss, and the stimuli were words used to label six concepts: velocity, acceleration, momentum, force, work and power. Multidimensional scaling procedures were applied to similarity ratings and measures of associative similarity. These procedures revealed that scaled estimates of psychological distance between concepts were correlated with the logical distance between concepts. This research bridges the gap between previous studies using words as stimuli and the research which is reported here.

Some additional research which bears upon the experiments contained in this report was done on the child's understanding of the concept of weight (Murray & Johnson, 1968; 1969). Here the independent variables for experimental study were based upon equations produced by the generator (e.g., the equation $F = \frac{Gm_1m_2}{D^2}$), and it

was found that psychologically relevant and irrelevant transformations for the concept of weight could be specified from such an analysis. From this research and that done with graduate students in physics, we conclude that our procedure has the potential for generating experimentally testable hypotheses as regards knowledge of concepts on various levels of scientific discourse, from initial stages of acquisition to advanced stages of scientific thinking.

Experimental Procedure

The experiments in this section of the report were designed to determine (a) how S's estimates of similarity among alpha-numeric expressions in physics correspond with features of the model used to generate them, and (b) how the structure of S's responses to alpha-numeric stimuli compares with the structure of their responses to other visual stimuli whose nominal structure also conforms to the form of the model in Fig. 3.

Experiment 1

Stimuli. Four stimulus sets were constructed. The first set (labeled "mxt") consisted of the eight expressions listed above which were generated from Fig. 3 by beginning at point (1) with the operand $\frac{dx}{dt}$ and performing the operations indicated by each axis.

To determine whether Ss' responses to expressions in the mxt set might be accounted for mainly in terms of syntactic features, a second stimulus set was constructed. Expressions in this set (labeled "$\alpha\beta\gamma$") were constructed by substituting α, β, and γ for m, x, t, respectively, in the expressions for the mxt set. Expressions in the $\alpha\beta\gamma$ set were identical to those in the first set in mathematical form, but were devoid of physical meaning.

To determine whether Ss' response to expressions in the mxt set might be dependent upon one particular syntactical form (mathematical representation), a third stimulus set (labeled "gm-cm") was constructed. To generate expressions for the gm-cm set, each of the eight expressions in the mxt set was written in terms of grams (gm), centimeters (cm), and seconds (sec), (e.g., $\frac{dx}{dt}$ was written as cm/sec). Expressions in the gm-cm set were the same semantically but different syntactically from expressions in the mxt set.

To determine whether Ss' responses on each of the above three stimulus sets might be accounted for solely in terms of an underlying transformational structure, a fourth stimulus set was constructed. This set (labeled "color") consisted of colored geometric forms varying on three binary dimensions: size, color, and shape. Although these dimensions do not directly correspond to the dimensions of the first three stimulus sets, they can be represented in the model in chapter 2 (Fig. 3) by arbitrarily assigning dimensions to axes. Motion in a positive direction on the p axis thus corresponds to a change in size from small to large. Motion in a positive direction on the q axis corresponds to a change in color from red to yellow, and motion in a positive direction on the r axis corresponds to a change in shape from square to triangle.

The stimuli in each set were photographed on 2" x 2" slides using a 35 mm Nikkormat camera with color film. A complete listing of the stimuli in each set is presented in Table 1.

Tasks. Two types of similarity judgments were obtained from each S to pairs of stimuli from each of the four sets. In the single pair comparisons task, each S was shown all 28 possible pairings of the eight stimulus items, one pair at a time. Ss were asked to rate the similarity of each pair on a scale from 1 (maximum similarity) to 11 (maximum dissimilarity). In the single pair comparisons, Ss were asked how similar the items appeared (for the αβγ set "as calculus expressions" and for the mxt and gm-cm sets, as "concepts in physics"). Stimuli were presented on a frosted backlit screen at a 10 sec. rate by two synchronized Kodak Carousel projectors connected to an automatic timer.

Immediately following the single comparisons tasks, Ss were asked to make double pair comparisons. For this task, stimuli on the three edges with a common vertex in Fig. 3 (chapter 2) were compared with one another in pairs (a modification of the method of triads). For example, for vertex #5, the three pairs compared would be 1-5 with 5-6; 1-5 with 5-8; and 5-6 with 5-8. Similarly, the three face diagonals with a common vertex were compared with one another in pairs. Again for vertex #5, the three pairs compared would be 5-2 with 5-4; 5-2 with 5-7; and 5-4 with 5-7. Thus, there were six pairs compared for each of the eight vertices, or 48 double pairs. For the doubles comparisons, Ss were asked to choose which pair of items appeared most similar (for αβγ, "as calculus expressions," for mxt and gm-cm, "as concepts in physics"). Stimuli for the double pair comparison task were presented on a backlit screen at a 15 sec. rate.

Table 2

Mean Dissimilarity Judgments and Standard Deviations for Each Stimulus Set

Category	Pair[a]	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
p edge	1-2	1.94	0.82	2.94	1.30
	3-4	2.18	1.18	2.88	1.75
	5-6 **	2.71	1.65	3.06	2.22
	7-8 *	1.94	0.66	2.41	0.79
q edge	5-1 *	7.47	2.85	5.94	2.38
	6-2 *	6.41	2.42	5.82	2.98
	7-3	6.71	2.82	5.59	3.00
	8-4 *	6.76	2.84	5.76	2.22
r edge	1-4	5.24	2.51	3.94	1.95
	2-3	5.88	2.29	4.71	1.90
	5-8 **	6.12	2.85	5.71	2.26
	6-7 *	6.41	3.12	5.65	1.77
pq face diagonal	1-6 *	6.47	2.48	6.24	2.56
	2-5 *	6.41	2.48	6.88	2.00
	3-8 *	7.18	3.05	6.06	2.49
	4-7	7.82	2.32	5.65	2.32
pr face diagonal	3-1	7.41	1.73	6.12	2.29
	4-2	7.29	2.57	5.65	1.66
	7-5 *	6.82	2.98	7.18	2.40
	8-6 **	7.06	2.66	7.59	1.62
qr face diagonal	1-8 *	6.76	3.88	6.12	2.95
	2-7	6.76	3.21	5.41	2.67
	3-6 *	8.47	2.65	8.00	2.29
	4-5 *	8.24	2.28	7.94	2.82
pqr grand diagonal	5-3 *	9.18	1.98	7.76	1.95
	6-4 *	9.35	1.22	7.88	1.96
	7-1	7.29	3.31	7.53	2.85
	8-2 *	6.65	3.52	5.88	3.06

[a] * indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

** indicates both members of the pair a e unlabeled.

Table 2

(cont'd)

Category	Pair[a]	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
	1-2	4.18	2.27	3.12	1.54
p	3-4	2.59	2.40	2.76	1.20
edge	5-6 **	5.12	2.60	2.94	1.44
	7-8 *	4.06	2.01	3.23	1.60
	5-1 *	6.35	2.64	2.76	1.09
q	6-2 *	5.88	2.64	3.59	1.62
edge	7-3	5.06	2.51	2.71	0.92
	8-4 *	4.94	2.58	2.88	1.11
	1-4	3.24	1.56	5.12	2.42
r	2-3	4.59	2.29	5.24	2.20
edge	5-8 **	6.59	3.16	4.82	2.21
	6-7 *	4.29	2.09	4.94	2.14
	1-6 *	7.94	2.41	4.12	1.11
pq	2-5 *	7.41	2.40	4.82	1.55
face	3-8 *	6.65	2.34	4.94	1.64
diagonal	4-7	6.47	2.24	4.29	1.76
	3-1	6.88	2.47	7.00	2.06
pr	4-2	6.59	2.50	6.18	2.33
face	7-5 *	7.00	2.78	7.24	1.95
diagonal	8-6 **	8.06	1.48	6.35	2.25
	1-8 *	3.12	1.62	7.12	2.38
qr	2-7	6.12	2.45	7.24	2.44
face	3-6 *	7.47	1.70	8.06	1.82
diagonal	4-5 *	7.71	2.52	7.12	2.32
	5-3 *	7.88	2.32	9.58	1.17
pqr	6-4 *	8.35	2.00	9.29	1.53
grand	7-1	6.00	2.45	9.29	1.53
diagonal	8-2 *	6.82	2.38	9.35	1.32

[a] *indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

**indicates both members of the pair are unlabeled.

diagonal judgments. The complete ordering relation was not satisfied in any stimulus set. We assume, however, there are not sufficient grounds for rejecting the assumption of interdimensional additivity in our data (cf. Beals, Krantz & Tversky, 1969; Tversky & Krantz, 1969). Therefore, the mean judgments for all stimulus sets were submitted to both metric (linear) and non-metric multidimensional scaling. Kruskal's (Version 5M) program was used for this purpose and scaling solutions were obtained by utilizing the structure embodied in Fig. 3 as a starting configuration.

The results of the nonmetric scaling in Euclidean distance showed that the eight stimulus points could be represented in three dimensions with stresses of: .001, .005, .018, and .000 for αβγ, mxt, gm-cm and color respectively. In the metric case, comparable stresses were .076, .087, .084, and .045. The resulting configurations (both metric and nonmetric) for the color set were rectangular polyhedrons, whereas the configurations for the other three sets were interpreted as distortions of the cube embodied in Fig. 3.(chap. 2). The configurations for the metric scaling of the mxt set is presented in Fig. 1.

An index of the agreement between scaled configurations for comparable stimulus sets was determined by computing both product moment and rank order correlations across the 28 interpoint distances in each metric solution. These correlations were .89 (.81), .53 (.76), and .75 (.84) for αβγ with mxt, αβγ with gm-cm, and mxt with gm-cm respectively. Comparable correlations based upon mean dissimilarity judgments were .91 (.77), .68 (.67), and .76 (.81).

The results of the scaling analyses indicated that the configuration for the color set was the most regular of all three dimensional solutions. This is reasonable under the assumption that the structure of the color set was the simplest of all four sets. Accordingly, Ss' responses to color stimulus pairs can be accounted for by transformational differences among the pairs, i.e., pairs represented on edges differed by one transformation, pairs represented on face diagonals differed by two transformations, and pairs represented on grand diagonals differed by three transformations.

Of course, this same structure underlies each of the other three sets. In the αβγ and mxt sets, the syntactic structure was transformational except that differences between members of pairs can be represented by the operations of calculus. And, of the four stimulus sets, the configurations for αβγ and mxt were most alike. In the gm-cm set, the syntactic structure was also transformational, but differences between members of a pair

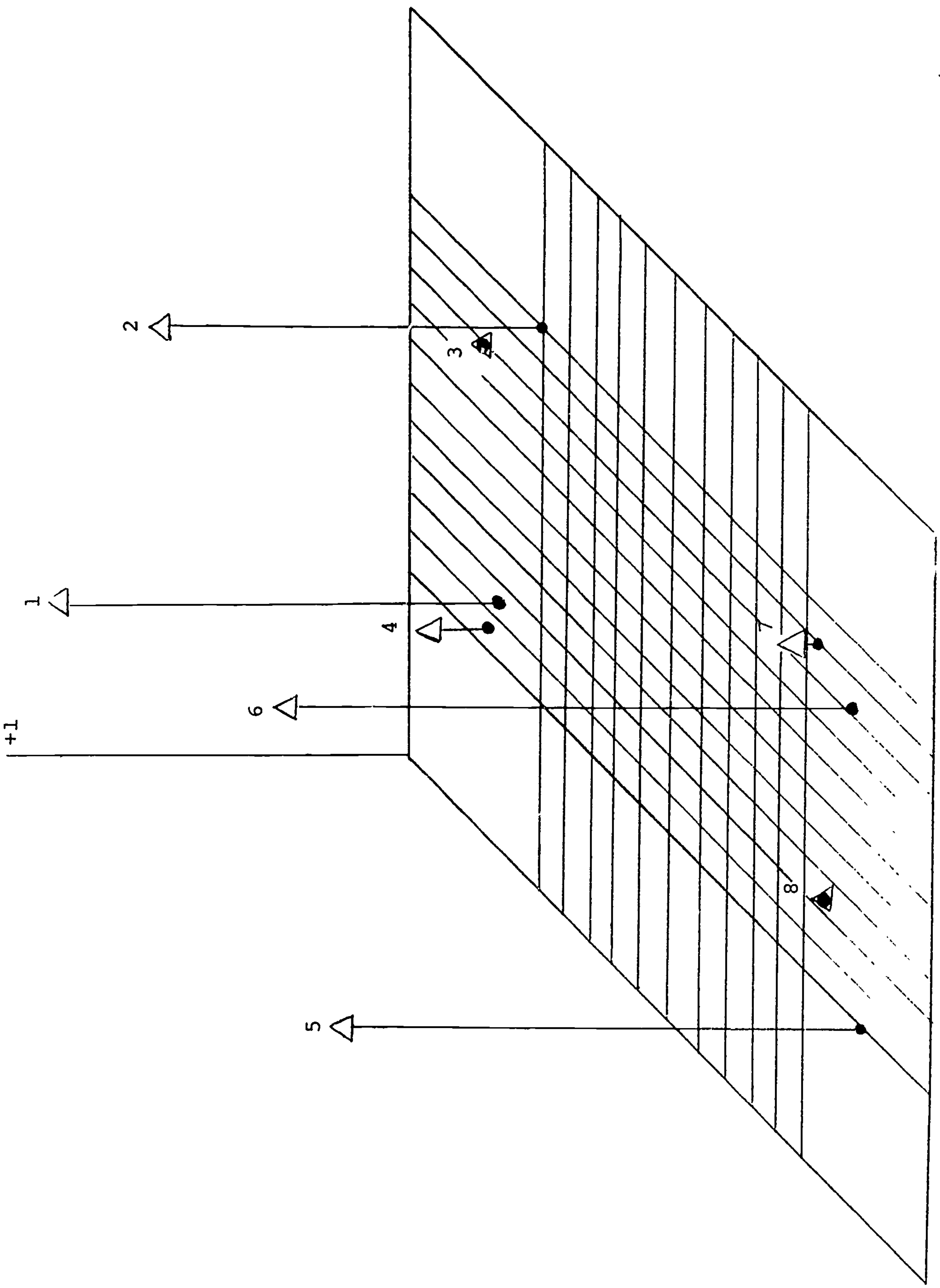


Fig. 1. Metric scaling of mxt dissimilarity judgments.

can be represented by the operations of ordinary algebra. While the configurations for mxt and gm-cm were similar, the discrepancies between them may be due to these syntactic differences.

The double-pair comparison task provides two additional types of information about the structure of each stimulus set: (1) it provides a means of verifying the differences between stimuli reflected in Ss' similarity judgments on the single pair comparison task and (2) it allows one to infer whether Ss were making their judgments in the singles task on the basis of some strategy such as counting transformations.

For each pair of pairs contrasted in the doubles task, S was consistent with his singles task performance if he chose whichever pair he rated as more similar when the pairs were presented individually. For example, on the singles task, S might have assigned a rating of 2 to the pair 5-1 and a rating of 3 to the pair 5-6 (see Fig. 3, Chap. 2). When presented with both pairs on the doubles task, i.e., 5-1 with 5-6, S should choose pair 5-1 as more similar than 5-6.

Table 3 presents the percent agreement (A) and percent disagreement (D) between the double and single pair comparisons. The percentages in columns labeled I (identical) refer to the situation where Ss rated both pairs as equally similar in the singles task. Note that only those comparisons which were made in the doubles task are included in this table, i.e., pairs of items differing on the same number of transformations.

The data in Table 3 indicate less disagreement between responses in single and double pair comparisons for the $\alpha\beta\gamma$ and color stimulus sets, than for the mxt and gm-cm stimulus sets. Since the maximum disagreement between performance on the two tasks is only 30 per cent we conclude that Ss generally did not make their judgments by counting transformations. This conclusion is supported by the percentage of I judgments. If Ss were counting transformations, the frequency of I judgments should have been close to 100%. Reference to Table 3 shows, however, that the percentage of I judgments was relatively low (15-30%) for all stimulus sets.

On the basis of the first experiment, it seems that Ss' responses reflect interrelations among the features used to construct each stimulus set. Furthermore, while both the ordering and magnitude of similarity judgments were generally consistent with these interrelations, Ss' responses seemed to be relatively free of the influence of a counting strategy.

Table 3

Contrast Between Double and Single Pair Comparisons for Each Task[a]

Stimulus pairs	αβγ,			mxt			gm-cm			color		
	A	D	I	A	D	I	A	D	I	A	D	I
Edges	75	10	15	58	24	19	43	30	27	61	12	27
Diagonals	55	19	26	56	26	19	47	31	21	54	16	30

[a]A indicates percent agreement, D indicates percent disagreement, I indicates identical judgments.

Table 4

Intertrial Correlations for Response (R) and Response Latency (RL) in Each Stimulus Set

Trials	αβγ		mxt		gm-cm		color	
	R	RL	R	RL	R	RL	R	RL
1-2	.99	.59	.98	.23	.97	.50	.99	.21
1-3	.99	.61	.98	.59	.96	.52	.99	.28
2-3	.99	.86	.99	.65	.98	.46	.99	.35

It is also true, as shown in Table 3, that the greatest disagreement between the singles and doubles task performance occurred for those stimulus sets with semantic (physical) content, namely mxt and gm-cm. It is possible, therefore, that Ss responded in the single comparison task to semantic as well as syntactic features in the two physics sets, but this was not evident in either the magnitude or ordering of their judgments.

Experiment 2 was constructed in an attempt to further index the influence of semantic features upon Ss' judgments to stimulus pairs in the mxt and gm-cm sets, and also to replicate the findings of experiment 1.

Experiment 2

Method

In experiment 2, each S was given three trials on a single pair comparison task with each of the four stimulus sets used in experiment 1. No doubles task was employed. As in experiment 1, Ss were instructed to determine the degree of similarity between the members of a stimulus pair. Response latency was collected for each stimulus pair.

Stimuli were presented upon a frosted backlit screen with unlimited exposure (i.e., until S responded) at a 5 second inter-stimulus interval. S responded by pressing buttons, numbered 1 to 11, corresponding to his judgment of the dissimilarity of the stimulus pairs. Response and response latency (RL) were automatically recorded.

Each of twenty-four Ss received all four stimulus sets. To control for order effects between sets, each S received one of 24 possible orders such that each set appeared in each ordinal position with equal frequency.

To control for order effects within sets, separate random orders of stimuli were constructed for each trial. The ultimate sequence of stimuli was different for each S.

Twenty-four graduate physics students from the University of Minnesota were tested individually and paid six dollars for participating in the two hour experiment. No S had participated in experiment 1.

Results and Discussion

The comparability of Ss' responses over trials was examined by correlating mean judgments and mean RL between pairs of trials in each of the four stimulus sets (see Table 4).

Correlations for the judgment data were uniformly high indicating that mean judgments in each of the four stimulus sets were remarkably alike from trial to trial. Correlations for the RL data, while somewhat lower than for judgment data, were generally higher between trials 2 and 3 than between trials 1 and 2. Subsequent analyses were performed on trial 3 for both judgments and RL since it seemed reasonable to assume that the performance of individual Ss was most stable on the last trial of the task.

As in experiment 1, means and standard deviations for Ss' judgments were computed for each stimulus pair. Trial 3 data for all four stimulus sets are presented in Table 5.

Inspection of the standard deviations indicates a range of variability within each stimulus set comparable to that found in experiment 1. This, plus an examination of the individual S data in experiment 2, suggests that even though performance becomes stable with practice, there remain differences between Ss which are reflected in the numbers they chose to indicate their judgments.

Mean judgments within any one category in Table 5 appear more alike than in experiment 1. And the ordering of face diagonals was better predicted by the ordering of edges than in experiment 1. In fact, with one exception, a perfect ordering was achieved in the color set.

Inspection of the means for the mxt and gm-cm sets indicates that pairs containing two labeled expressions were judged more similar than pairs in the same category with one or more unlabeled expressions. Since an expression's labelability as a concept in physics must depend upon the presence or absence of semantic features, this finding indicates that, with practice, the magnitude of Ss' judgments reflects the meaning of expressions. Consequently, even though Ss' judgments generally correspond with the (idealized) syntactic structure in Fig. 3 (Chap. 2), they can also be interpreted as a distortion in this structure based upon the manner in which combinations of operands have been identified with concepts in physics.

Table 5

Mean Dissimilarity Judgments and Standard Deviations on Trial 3 for Each Stimulus Set

Category	Pair[a]	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
p edge	1-2	2.21	1.03	2.75	1.56
	3-4	2.17	1.02	2.75	1.47
	5-6 **	2.63	1.47	2.83	1.41
	7-8 *	2.71	1.74	3.21	1.85
q edge	5-1 *	5.29	2.40	5.08	2.13
	6-2 *	5.17	2.05	5.17	2.13
	7-3	5.29	2.29	4.46	1.58
	8-4 *	5.25	2.09	4.50	1.66
r edge	1-4	4.67	2.02	4.63	1.80
	2-3	5.00	2.00	4.63	1.65
	5-8 **	5.38	1.98	6.04	1.63
	6-7 *	6.04	2.06	5.96	2.09
pq face diagonal	1-6 *	6.04	2.14	6.25	2.04
	2-5 *	6.21	1.91	6.67	2.12
	3-8 *	6.21	2.38	6.42	1.81
	4-7	6.33	2.55	6.04	1.60
pr face diagonal	3-1	6.25	2.30	6.17	2.05
	4-2	5.63	2.17	6.08	1.92
	7-5 *	6.67	2.31	7.88	1.64
	8-6 **	6.67	2.08	7.33	1.59
qr face diagonal	1-8 *	6.08	2.47	6.29	2.54
	2-7	6.54	2.21	6.42	2.41
	3-6 *	8.08	2.02	7.71	1.64
	4-5 *	7.83	2.07	7.63	1.85
pqr grand diagonal	5-3 *	8.33	1.72	8.67	1.61
	6-4 *	8.63	1.81	8.21	1.94
	7-1	7.25	3.73	7.21	2.38
	8-2 *	7.38	2.53	7.22	1.72

a *indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

**indicates both members of the pair are unlabeled.

Table 5

(cont'd)

Category	Pair[a]		gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
	1-2		3.79	1.91	3.21	1.23
p	3-4		2.75	1.62	3.21	1.51
edge	5-6	**	4.21	2.08	3.13	1.30
	7-8	*	3.83	1.88	3.16	1.56
	5-1	*	5.58	2.57	3.04	1.46
q	6-2	*	5.67	2.45	3.00	1.50
edge	7-3		4.92	2.25	3.13	1.49
	8-4	*	5.21	2.57	3.08	1.35
	1-4		4.12	1.58	5.75	2.13
r	2-3		4.71	2.05	5.58	2.25
edge	5-8	**	5.75	2.92	5.67	2.40
	6-7	*	6.38	2.67	5.46	2.26
	1-6	*	7.50	2.04	4.88	1.76
pq	2-5	*	7.92	1.96	4.63	1.65
face	3-8	*	7.04	2.16	4.92	1.83
diagonal	4-7		6.83	2.15	4.88	1.61
	3-1		6.83	2.21	6.79	2.21
pr	4-2		6.33	2.15	6.88	2.17
face	7-5	*	7.83	2.27	7.04	2.31
diagonal	8-6	**	7.58	2.15	6.83	2.19
	1-8	*	4.33	1.67	7.29	2.26
qr	2-7		5.67	2.29	6.29	2.22
face	3-6	*	7.33	2.13	7.00	2.35
diagonal	4-5	*	7.50	2.55	5.13	2.10
	5-3	*	8.42	1.75	8.92	2.00
pqr	6-4	*	8.29	1.99	8.92	2.21
grand	7-1		6.75	2.17	8.88	2.05
diagonal	8-2	*	7.58	1.94	9.03	1.94

[a] *indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

**indicates both members of the pair are unlabeled.

Mean judgments in different stimulus sets were examined by submitting the means in Table 5 to both metric and nonmetric multidimensional scaling analysis. The results of nonmetric scaling showed that the eight stimulus points could be represented in three dimensions (Euclidean distance) with stresses of .000, .000, .039, and .000 for αβγ, mxt, gm-cm, and color, respectively. In the metric case, the comparable stresses were .037, .045, .081, and .030. For each stimulus set, the metric and nonmetric solutions were highly similar and less distorted than comparable configurations for experiment 1. The configuration for the metric scaling of the mxt set is presented in Figure 2.

As in experiment 1, the scaling solutions reflect the syntactic features built into each stimulus set. In the case of the αβγ and mxt sets, the features consisted of calculus operators (integration and differentiation), while in the case of the gm-cm set, the features were multiplication (including squaring), and division. For color, the features were transformations within a category, i.e., from small to large, from red to yellow, and from square to triangle.

Ss' judgments in different stimulus sets were compared by computing product moment (and rank order) correlations across the 28 interpoint distances in each metric solution. These correlations were .96 (.95), .87 (.92), and .92 (.94) for αβγ with mxt, αβγ with gm-cm, and mxt with gm-cm, respectively. Comparable correlations based upon mean judgments were .96 (.96), .86 (.87), and .90 (.90). These correlations imply that Ss responded much alike in the three stimulus sets. However, a closer examination of the means in Table 5 indicates that pairs in the p category for the αβγ set were uniformly judged more similar than comparable pairs for mxt and gm-cm sets.

Reference to Table 1 shows that in the αβγ set, the p edge represents multiplication by a constant (α). Evidently in making similarity judgments among expressions consisting solely of calculus operations, multiplication by a constant was relatively less important, dβ and d/dγ being the primary means used to rate dissimilarity. In the mxt and gm-cm sets, the p edge also represents multiplication by a constant (mass). In this case, however, the constant has physical meaning, and so entered into Ss' judgments. This finding can be considered further evidence for the conclusion that Ss were attending to semantic as well as syntactic features of stimulus pairs in the mxt and gm-cm sets.

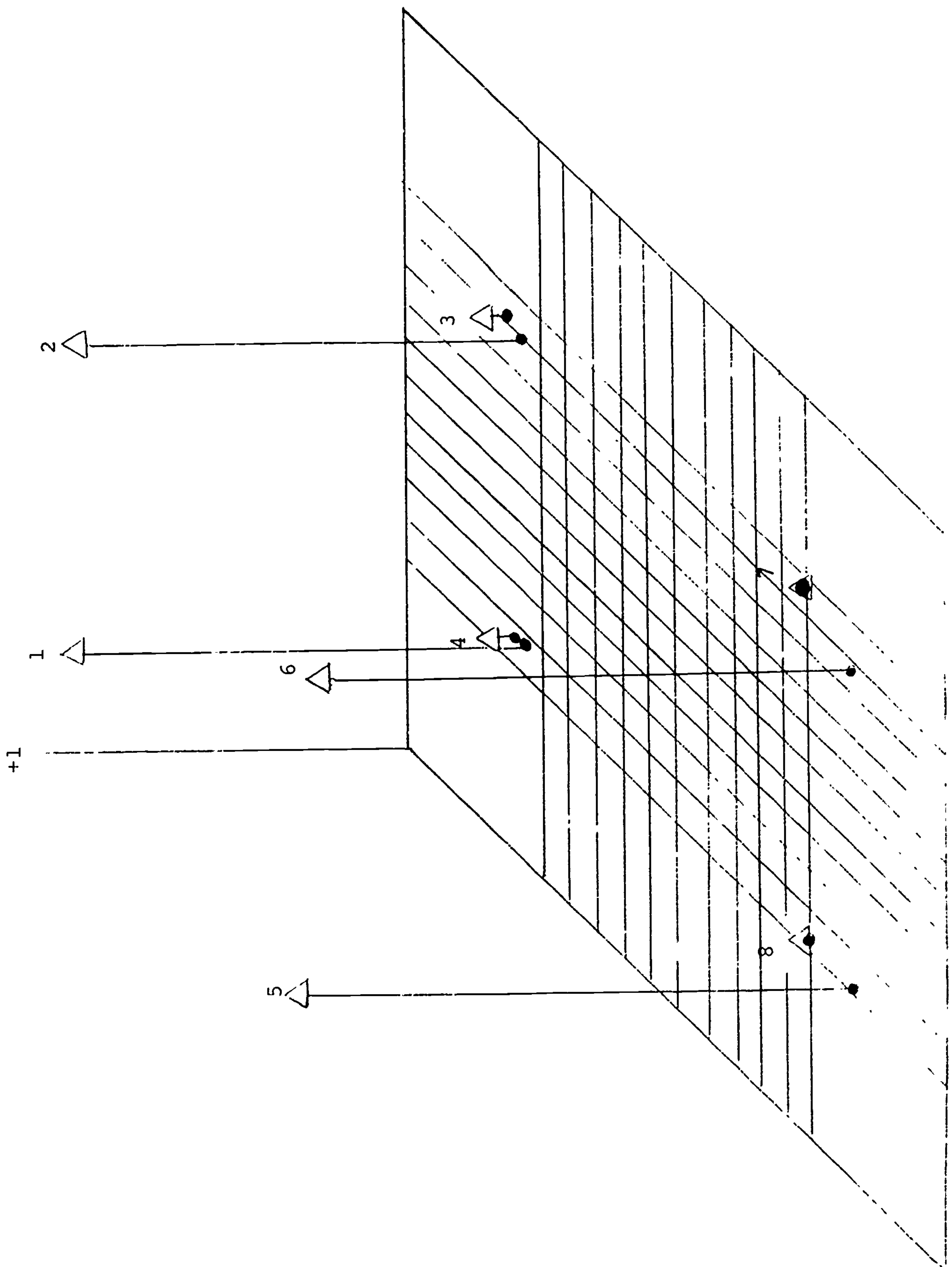


Fig. 2. Metric scaling of Trial 3 mxt dissimilarity judgments.

Latency data were next examined to provide additional information as to the nature of $\underline{S}$s responses in each stimulus set. Correlations were computed between dissimilarity judgments and RL for each trial. Although the correlations increased from trial 1 to trial 3, they were sufficiently low to indicate the RL should provide additional information as to how $\underline{S}$s responded in the judgment task. (By trial 3 these correlations were .64, .66, .36, and .24 for αβγ, mxt, gm-cm, and the color sets, respectively.)

Mean latencies and standard deviations were computed for $\underline{S}$s' judgments on trial 3 for each stimulus pair (see Table 6). Inspection of the standard deviations shows comparable between $\underline{S}$ variability for pairs in the mxt and gm-cm stimulus sets, with lower variability in the αβγ and color sets. Inspection of the means indicates the RL for αβγ pairs was generally less than RL for mxt and gm-cm pairs, while RL for color pairs was lowest of all four sets.

Product moment (and rank order) correlations were computed across mean RL for comparable stimulus pairs. These correlations, which were .79 (.63), .57 (.53), and .54 (.43) for αβγ with mxt, αβγ with gm-cm, and mxt with gm-cm, respectively, indicate that the mean latency of $\underline{S}$s' judgments was most similar for pairs with comparable syntactic features (i.e., mxt and αβγ). Inspection of means within the mxt and gm-cm stimulus sets reveals a trend for RL to be shorter for pairs involving two labeled expressions than for pairs containing at least one unlabeled expression.

For each stimulus set, response latencies were averaged across $\underline{S}$s, and across the 28 stimulus paris for each trial, to examine trends in RL with practice. The mean latencies for each trial are presented graphically in Fig. 3, which shows an asymptotic decrease in RL as a function of trials for all stimulus sets.

Findings based upon latency data lend modest support to the earlier conclusion that within the limitations of the stimulus sets, $\underline{S}$s responded to stimuli in each set in terms of features embodied in the model in Fig. 3 (Chap. 2). The influence of different types of features (semantic vs. syntactic) was not clear cut due to the small differences on means between sets and the considerable amount of variability among $\underline{S}$s for a given stimulus pair on any trial. If one assumes that repeated presentations stabilize $\underline{S}$s' performance over trials, the variability among $\underline{S}$s on trial 3 suggests individual differences worth pursuing.

Table 6

Mean Response Latencies and Standard Deviations on Trial 3 for Each Stimulus Set

Category	Pair[a]	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
	1-2	2.28	0.78	2.31	1.06
p	3-4	2.82	1.65	2.86	1.18
edge	5-6 **	3.28	1.10	3.62	1.11
	7-8 *	4.11	0.86	3.67	1.81
	5-1 *	3.08	1.24	3.30	1.66
q	6-2 *	3.72	1.55	4.02	2.80
edge	7-3	3.78	1.28	4.13	2.06
	8-4 *	3.57	1.78	3.54	2.04
	1-4	2.98	1.33	3.01	1.20
r	2-3	3.28	1.17	3.49	1.88
edge	5-8 **	3.63	1.57	5.00	3.04
	6-7 *	4.14	1.60	4.42	2.99
	1-6 *	8.76	1.54	3.70	1.32
pq	2-5 *	3.65	1.34	3.92	2.01
face	3-8 *	3.50	1.10	5.09	2.42
diagonal	4-7	4.08	1.80	4.20	1.85
	3-1	3.88	2.01	3.67	2.24
pr	4-2	3.43	1.44	3.12	1.10
face	7-5 *	4.23	1.80	4.19	1.93
diagonal	8-6 **	4.10	1.81	4.57	2.34
	1-8 *	3.78	1.48	4.15	2.85
qr	2-7	4.65	1.69	5.48	1.33
face	3-6 *	4.05	2.02	4.89	2.91
diagonal	4-5 *	3.45	1.08	3.85	2.20
	5-3 *	4.27	1.83	5.26	1.70
pqr	6-4 *	3.83	1.70	4.06	2.50
grand	7-1	4.21	1.99	4.24	2.21
diagonal	8-2 *	4.08	1.89	4.49	2.34

[a] *indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

**indicates both members of the pair are unlabeled.

Table 6

(cont'd)

Category	Pair[a]	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
p edge	1-2	3.19	1.42	2.35	0.89
	3-4	3.60	1.09	2.55	1.14
	5-6 ***	4.47	2.58	2.50	1.27
	7-8 *	3.84	1.47	2.23	0179
q edge	5-1 *	3.17	1.20	2.17	1.25
	6-2 *	4.22	2.20	2.27	1.27
	7-3	4.30	1.62	2.12	0.87
	8-4 *	3.72	1.64	2.09	0.85
r edge	1-4	3.81	1.91	2.84	1.45
	2-3	4.02	1.82	2.52	1.06
	5-8 **	3.71	1.74	2.84	1.47
	6-7 *	4.56	2.60	2.70	0.97
pq face diagonal	1-6 *	4.75	2.76	2.85	1.68
	2-5 *	4.30	2.35	2.94	1.42
	3-8 *	5.64	1.68	2.57	1.16
	4-7	4.30	1.76	2.70	1.22
pr face diagonal	3-1	4.92	2.76	2.51	1.18
	4-2	4.58	1.98	3.19	1.73
	7-5 *	4.47	2.03	2.36	0.86
	8-6 **	4.38	2.27	2.46	1.12
qr face diagonal	1-8 *	3.20	0.99	2.71	1.23
	2-7	6.63	1.90	3.19	1.91
	3-6 *	4.45	1.62	2.45	1.01
	4-5 *	3.90	2.56	2.74	1.58
pqr grand diagonal	5-3 *	4.01	2.20	2.47	2.23
	6-4 *	4.24	1.66	2.77	1.43
	7-1	5.39	2.65	2.04	0.73
	8-2 *	4.05	1.41	2.27	1.48

[a] *indicates one unlabeled expression in the pair (for the mxt and gm-cm set).

**indicates both members of the pair are unlabeled.

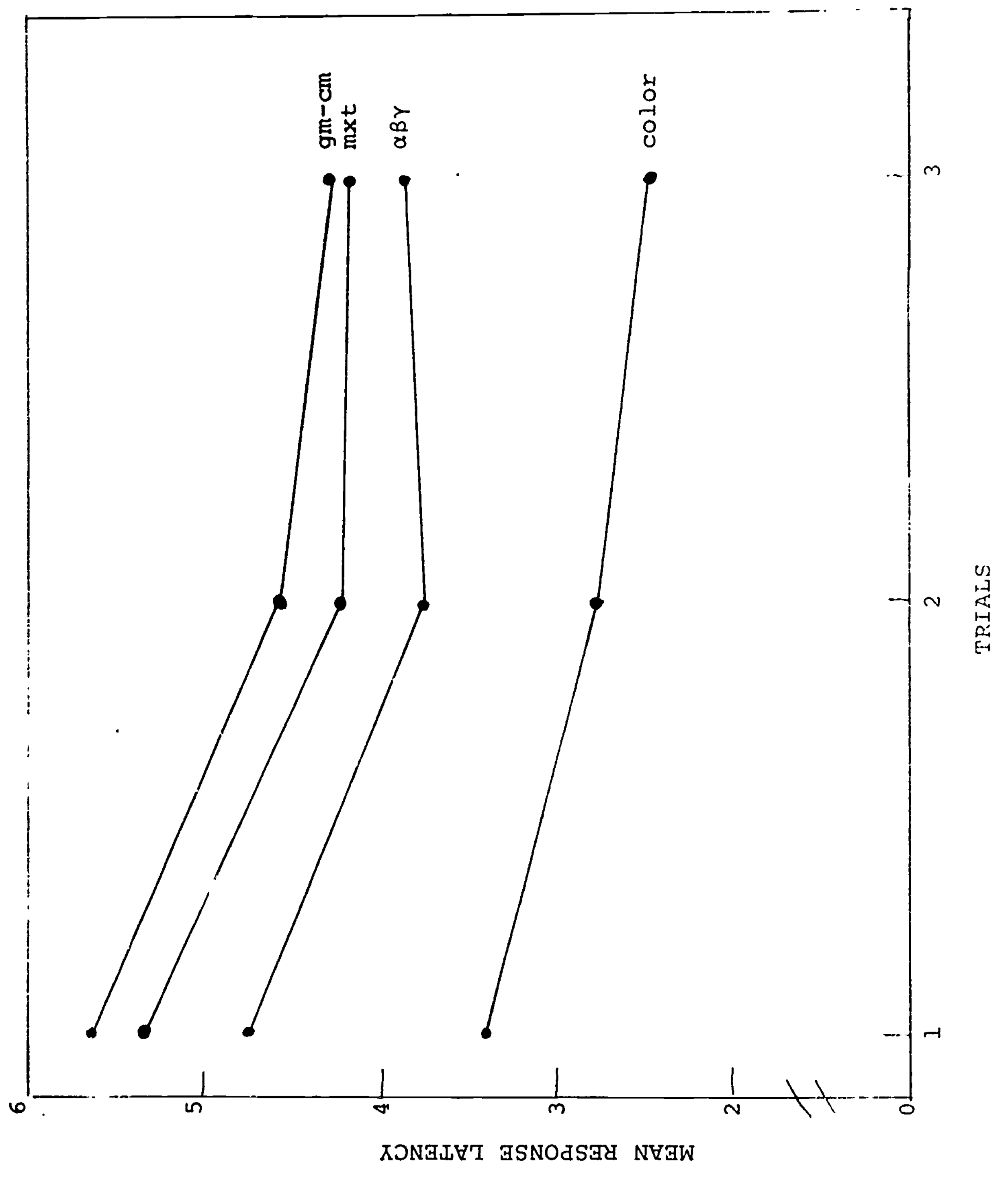


Fig. 3. Mean Response Latency by Trial for All Pairs in each Stimulus Set

We have assumed that distortions in the scaling configurations of experimental stimuli due to individual differences are not large (cf. Delbeke, 1968, p. 29; Horan, 1969). REcently available techniques in individual difference scaling (Carroll & Chang, 1971), or more simply procedures suggested by Tversky & Krantz (1969), should be employed as a check on our results.

We conclude from the performance of Ss in the two prototype experiments that a structure of knowledge in physics can be represented by the model in Fig. 3 (Chap. 2). More specifically, we suggest that knowledge of the syntactic features of mathematical expressions in physics is represented in the correspondence between Ss' responses and the general form of the model's structure. Knowledge of the semantic features of these expressions may then be represented in the correspondence between Ss' responses and small but regular distortions of this structure. Because the structure underlying knowledge of expressions corresponds to performance on the simpler color set, we speculate that the specialized knowledge acquired from instruction in science can be viewed as a complication or perturbation in already existing structures of some generality (Deese, 1969b).

In concluding this chapter it should be pointed out that our discussion of knowledge has been limited by the fact that the abstraction used to interpret Ss' performance is basically a syntactic structure. That is to say, Fig. 3 is a description of idealized relations among expressions in physics based upon syntactic features of the language in which these expressions are written. Accordingly, if one takes the view (as we do) that understanding occurs when content is assimilated to conceptual categories (Deese, 1969a), the knowledge accounted for here leads primarily to a syntactic (in this case, mathematical) understanding of content.

It is true, however, that the semantic features of expressions in the mxt and gm-cm sets can be interpreted by reference to Fig. 3, provided that external criteria are used to indicate which expressions have content labels and which do not. In this sense, the model is as good a description of semantics as it is of syntax. But semantics in science is more than interrelations among abstract concepts; it must also be a function of the manner in which these concepts apply to experience (perceptually, as well as cognitively - cf., Bohm, 1965, p. 226). Semantics in this latter (though still restricted) sense is embodied in the prototypical problems of science (Kuhn, 1971). Examples of these in the case of classical physics are the falling body, the harmonic oscillator and the pendulum.

WHile the expressions serving as stimuli in the present studies were relatively simple, it should be clear that more complex expressions can be generated from Fig. 3, simply by beginning with a more complex notation, i.e., one need not start with $\frac{dx}{dt}$; one could just as well start with $e^x \sin \theta t$. In general, the form of an expression generated for a particular point in the model depends upon where one starts, the expression one starts with, and the "path" taken to reach that point. There was no path dependence for the expressions used in the two experiments reported here because all integrations were assumed to be indefinite. Ordinarily, however, one integrates over limits in applying mathematical concepts to a real (or prototypical) physical situation. It is these limits in the form of constants of integration that tie abstract expressions such as $m\frac{d^2x}{dt^2}$ to reality.

The semantic meaning of symbol structures in a science such as physics is ultimately embodied in properties of natural systems such as initial and final conditions as well as symmetries and physical constants. An adequate psychological understanding of scientific knowledge requires that these properties be included in the abstractions we use to guide our study and experimentation.

References

Bar-Hillel, Y. Universal semantics and philosophy of language: Quandries and prospects. In Puhvel (Ed.), Substance and Function of Language. California: 1969.

Beals, R., Krantz, D. H., and Tversky, A. Foundations of multidimensional scaling. Psychological Review, 1968, 75, 127-142.

Bloomfield, L. Linguistic aspects of science. International Encyclopedia of Unified Science, 1939, 1, 1-59.

Bohm, D. Physics and perception. In D. Bohm (Ed.), The Special Theory of Relativity. New York: W. A. Benjamin, 1965.

Carnap, R. Philosophical Foundations of Physics. New York, 1966.

Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika, 1970, 35, 283-319.

Deese, J. Behavior and fact. American Psychologist, 1969a, 24, 515-522.

Deese, J. Conceptual categories in the study of content. In G. Gerbner (Ed.), The Analysis of Communication Content. New York: Wiley, 1969b.

Delbeke, L. Construction of Preference Spaces: An Investigation in the Applicability of Multidimensional Scaling Models. Louvain-Leuven, Belgium: The University of Louvain, 1968.

Fillenbaum, S. & Rapoport, A. Structures in the Subjective Lexicon. New York: Academic Press, 1971.

Fodor, J. A. Psychological Explanation: An Introduction to the Philosophy of Psychology. New York, Rnadom House, 1968.

Garner, W. E., Hake, H. W. & Eriksen, C. W. Operationism and the concept of perception. Psychological Review, 1956, 63, 149-158.

Henley, N. M. A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 176-184.

Horan, C. B. Multidimensional scaling: Combininb observations when individuals have different perceptual structures. Psychometrika, 1969, 34, 139-165.

Johnson, P. E. An experimental analysis of written instruction. Verbal Learning Research and the Technology of Written Instruction. (E. Z. Rothkopf and P. E. Johnson, Eds.) Rand McNally, 1969b.

Johnson, P. E. Associative meaning of concepts in physics. Journal of Educational Psychology, 1964, 55, 84-88.

Johnson, P. E. On the communication of concepts of science. Journal of Educational Psychology, 1969, 60, 32-40.

Johnson, P. E. Some psychological aspects of subject matter structure. Journal of Educational Psychology, 1967, 58, 75-83.

Johnson, P. E. Word relatedness and problem solving in high school physics. Journal of Educational Psychology, 1965, 56, 217-224.

Johnson, P. E., Cox, D. L., & Curran, T. E. Psychological reality of physical concepts. Psychonomic Science, 1970, 19, 245-247.

Kuhn, T. Second thoughts on paradigms. In T. Kuhn (Ed.), Structure of Scientific Revolutions.

Miller, R. G., Jr., Simultaneous Statistical Inference. McGraw-Hill, Inc., 1966.

Murray, F. B. & Johnson, P. E. Reversibility in non-conservation of weight. Psychonomic Science (in press).

Murray, F. B. & Johnson, P. E. A model for relevant and irrelevant transformations in children's concept of weight. Hausdorff, H. (ed.) A.E.R.A. Paper Abstracts, 1968 Annual Meeting, 308.

Newell, A. & Simon, H. Human Problem Solving. New Jersey: Prentice-Hall, 1972.

Shepard, R. N. & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology, 1970, 1, 1-17.

Simon, H. & Newell, A. Human problem solving: The state of the theory in 1970. American Psychologist, 1971, 26, 145-159.

Tisza, L. The logical structure of physics. Synthese, 1962, 14, 110-131.

Tversky, A. & Krantz, D. Similarity of schematic faces: A test of interdimensional additivity. Perception & Psychophysics, 1969, 5, 124-128.

Chapter 4

Communication Task

The experiment reported in this chapter was designed to explore how the structure of Ss responses in a judgment task can be used to interpret discrimination among stimuli in a task designed to embody some of the variables of scientific communication. The experiment allows for a test of how communication tasks can be constructed in terms of encoding and decoding operations, as well as examining the usefulness of similarity judgments as indices of the cognitive structure of subject matter. In the encoding operations, S must choose a set of n stimuli such that this set seems odd when contrasted with an E designated stimulus. In the decoding operations, Ss are given sets of encoded stimuli and asked to select the "odd" stimulus.

The study is based in part upon the work of Isaac (1968) who investigated the perceived similarity among a set of schematic faces. A point of departure from Isaac is that we have a formal model from which to generate message sets and against which to measure Ss' performances. That formal model specifies certain geometric relations which logically obtain among the stimuli and obviates reliance on a set of stimuli with uncertain relations.

As mentioned above, the communications tasks consists of encoding and decoding operations (Rosenberg and Cohen, 1966). Assuming that different Ss have different Weltanschauungen and that this difference in perception is reflected in intransigencies in communicating (Runkel, 1956), one might suspect cognitive differences between Ss. If, as Neisser (1966) points out (and to which most of us tacitly agree), perception is a matter of construction (or reconstruction), then the two stage theory of Rosenberg and Cohen offers a vehicle for the construction of communication tasks in terms of encoding (construction) and decoding (re-construction and comparison) operations. These tasks are further valuable in assessing the utility of dissimilarity judgments as indices of perceptual or cognitive structure (or vice versa), in that sense serving as converging operations in delimiting possible characterizations of a model of perceptual structure.

The encoding operations simply consist of presenting Ss with a set of stimuli from which E selects one and designates it as a target. Ss must choose a number (n) of other stimuli (contrast sets) from the remaining set such that they make the target seem "odd" or "inappropriate" in the new context (n + target). Each of the stimuli in the set is systematically used as the target stimulus and, in each case, the S's task is to find the set of n stimuli which make the target "odd".

In the decoding operation, Ss are given sets of stimuli (encoded or message sets) and are asked to select the "odd" stimulus in each set.

Ss must serve in the experiment upon two occasions. In the first session, Ss produce dissimilarity judgments to pairs of stimuli and, immediately thereafter, perform the encoding task. At a second session, Ss perform the decoding task.

Data are subjected to comparisons of

1. dj vs. encoding for each S
2. dj vs. decoding for each S
3. encoding vs. decoding
4. model vs. encoding
5. model vs. decoding

Comparisons to be made for each S also include performance according to whether S-defined sets are "own" (or done by him), "Other" (or encoded by other Ss), predicted sets (not encoded but consistent with djs), or E-defined sets (derived from the formal model).

In addition to whether Ss' responses on the communications tasks are predicted by their (dis)similarity judgments or the logical model, they may be found to be consistent with the following:

a. a distinctive feature analysis of the stimuli
b. physics "meaning" (e.g., do S's dichotomize) according to kinematics and dynamics, etc.)
c. labelability of expressions as physical concepts or constructs.

Method

Ss were 24 graduate physics students from the University of Minnesota. Ss were tested individually and were paid nine dollars for participating in a 3 hour experiment.

Three experimental procedures were conducted in two sessions. The first session consisted of dissimilarities judgment tasks and an encoding task. The second session consisted of a decoding task.

Session I -- Dissimilarity Judgment

Dissimilarity judgments were obtained from pairs of stimuli from the mxt set used in the two experiments described in chapter 3.

Encoding task.

Immediately after a S completed the dissimilarity judgments, he performed the encoding task. Materials consisted of the 8 mxt expressions printed one each on 2 1/2 x 3 inch cards. Recall that these expressions are concepts identified in a logical abstract cube in Fig.3(chap 2).

E shuffled the stimulus cards and presented them in a haphazard array, approximately rectangular, for each trial. One expression was selected by E as a "target" and placed to S's left of the remaining 7 cards. According to the target, S selected 3 of the remaining stimuli which made that target seem "odd" or inappropriate in their context. Thus, S constructs or encodes a message set of 4 expressions, one being odd.

S was told that the optimal strategy would be to choose contrast sets such that the 3 expressions were all more similar to each other than any was to the target. S was allowed to manipulate the materials as he desired and identified his selections by handing the contrast set to E. No time limit was specified.

Each of the 8 expressions served in turn as a target according to a pre-arranged random order. Thus, each S produced 8 message sets.

Of the original Ss, 17 completed the encoding task and 14 returned to complete the decoding task.

Session II -- Decoding Task

Fourteen Ss completed this portion of the experiment, all within 3 to 7 days after the first session.

Each S was given 67 mxt message sets, one at a time, and was instructed to choose the expression which seemed most odd in the set of 4. S was reminded of the previous Encoding Task and was told that he was to perform essentailly the reverse operation.

The expressions were typed and drawn on 4 x 5 inch cards and were placed in a loose leaf or ring binder. Each card contained one message set, i.e., 4 expressions

in a square array. There were four permutations of the array for each message set and four random orders. Thus, there were 16 permuted array x order stimulus sets.

As originally conceived, decoding sets were to be constructed on the bases of 4 paradigms: (a) S-defined ("own") sets, encoded by S himself, (b) "other" defined sets, encoded by others than the decoder, (c) predicted sets or sets constructed by E from the decoder's own dissimilarity judgments where such sets were not the same as (a) or (b), and (d) E-defined sets or sets of interest to E because of the relational characteristics of expressions in the abstract or formal model (cube). Based upon results of an earlier pilot study, the E-defined sets were constructed to reflect: (a) faces: 4 expressions constitute the vertices of any one face of the cube where each pair is assumedly $1 \leq St \leq 2$ units of distance apart; (b) 3 nodes of one face and one node connected thereto via an edge (i.e., relational positions on the cube describable by 3 edges, 2 diagonals, 1 super-diagonal); (c) expressions of 3 vertices connected by an edge to a common vertex and (d) 3 nodes of one face connected to one node of another face. The latter (d) are expected to obtain also from Ss' dissimilarity judgments.

Due to the results of the pilot study and because of convenience in preparing the stimulus materials, the E-defined sets dictated the selection of the 62 message sets presented to each S. In only 4 instances did Ss encode sets which were not in the 62 sets constructed by E-one of the 4 Ss involved did not participate in the decoding task. In a sense, then, all message sets were E-defined but contained, own, other, and predicted sets. The remaining 8 possible sets $\binom{8}{4,4}$ were excluded because of pilot study results.

Results

Dissimilarity Judgments:

The results of the dissimilarity judgments confirm the findings reported in chapter 3. Suffice it to say here that, for the mxt task, multidimensionally scaled data were nearly a duplicate of the model shown in Fig. 3. Individual S responses are reported in Table 1. It is these observations with which encoding and decoding tasks are compared for optimality.

Table 1

Individual Subject Responses on Trial 3

mxt Task

St \ S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1-2	2	2	4	5	2	8	3	6	3	2	2	2	1	2	2	2	2
3-4	2	2	4	6	2	7	3	4	5	2	1	2	1	2	2	2	2
5-6	2	2	4	6	3	7	3	4	3	2	2	2	1	3	2	2	4
7-8	2	2	4	6	2	8	3	4	5	2	2	2	1	2	2	3	8
1-5	3	3	5	4	7	7	5	4	3	5	5	2	3	9	6	6	6
2-6	3	2	5	4	8	7	6	4	3	7	6	4	4	3	6	5	5
3-7	3	2	5	3	4	7	6	4	4	5	2	4	3	3	7	6	5
4-8	3	2	5	4	4	7	4	5	3	6	3	4	2	3	6	6	7
1-4	3	1	5	6	3	7	6	7	3	4	2	3	4	3	6	5	6
2-3	3	4	5	6	3	8	6	7	3	3	2	3	4	3	5	5	4
5-8	5	7	5	5	5	7	5	7	5	5	7	7	4	8	5	6	6
6-7	3	3	5	7	7	10	6	7	4	7	5	4	4	7	6	6	6
1-6	4	5	8	5	8	10	6	4	4	6	4	7	6	5	6	7	8
2-5	4	4	8	6	7	11	8	6	5	5	5	7	6	10	5	7	7
3-8	3	6	9	7	4	9	9	5	5	7	3	7	6	4	7	7	9
4-7	4	4	8	4	4	9	7	6	6	7	3	7	5	5	7	7	7
1-3	4	5	5	8	4	8	9	8	5	7	4	4	8	3	6	5	10
2-4	5	4	8	8	3	11	7	8	5	5	3	5	7	4	7	6	8
5-7	5	8	9	8	7	10	8	9	8	7	5	10	8	10	6	7	8
6-8	6	9	8	7	7	11	7	7	6	6	5	5	8	6	6	7	9
1-8	6	9	5	4	6	10	2	8	6	7	3	5	6	7	2	5	5
2-7	4	7	9	7	4	11	7	9	7	8	4	7	8	4	2	4	4
3-6	6	7	7	6	7	10	8	8	7	8	7	7	9	7	8	10	11
4-5	5	9	6	8	8	9	9	8	4	6	6	6	8	9	8	10	10
3-5	6	9	9	7	7	10	9	8	7	8	6	10	10	9	8	11	11
4-6	6	7	8	8	8	10	8	7	6	7	3	10	9	9	8	11	11
1-7	6	11	9	7	4	11	5	9	7	10	4	10	7	3	4	5	5
2-8	6	7	9	6	5	10	6	8	6	8	6	10	7	6	6	6	8

Encoding Task:

Tests of optimal behavior (target -- contrast sets predicted by dissimilarity judgments) are made by examining the proportion of responses summed across all Ss and all targets falling into optimal (O), non-optimal (N) and Toss-up (T) categories.

Optimality was determined by rank sum test comparing the three target-contrast dissimilarities with the contrast-contrast dissimilarities. In order to be considered as O, the minimal stimulus rank sum must be 9 or less and at least 1.0 rank sum units from the next nearest rank sum.

The row marginals of Table 2 demonstrate that the proportion of O, N, and T selections of contrast sets are .639 (87/136), .272 (37/136), and .088 (12/136), respectively. Assuming the criteria established above, the probability of an expression having an O rank sum is .19 if there are no ties, thus predicting an expected number of 1.52 optimal sets of 8. The observed range for individual Ss was 2 to 8 (corresponding to proportions of .250 to 1.000). Assuming the same criteria but summing across Ss for a given target, the observed proportion of O sets ranged from .352 (6/17) for target 2 ($m\frac{dx}{dt}$) to .823 (14/17) for target 5 ($\int(\frac{dx}{dt})dx$).

Session II -- Decoding

Decoding Accuracy:

Accuracy of decoding is evaluated with respect to accuracy of decoding message sets encoded by oneself (own), predicted sets constructed by E based upon Ss' dissimilarity judgments but not encoded by S, and E-defined sets. The latter category allows for the generation of all predicted sets of interest.

The proportion of own sets correctly decoded is .725 (79/109). Three sets were encoded which were not contained in the decoding task.

Table 2

Proportion of Optimal, Non-optimal, and Toss-up Choices for Encoding Task

Sum of choices per S: S	O	NO	TU	Target Expressions: 1	2	3	4	5	6	7	8	p(O)	p(N)	p(T)
13	8	-	-	O	O	O	O	O	O	O	O	1.000	-	-
1	6	1	1	O	O	O	O	T	N	O	O	.750	.125	.125
2	7	1	-	O	N	O	O	O	O	O	O	.875	.125	-
3	4	3	1	O	N	O	N	O	N	T	O	.500	.375	.125
4	4	2	2	N	T	O	O	O	O	N	T	.500	.250	.250
17	4	4	-	N	N	O	O	O	O	N	N	.500	.500	-
6	4	3	1	T	O	N	N	N	O	O	O	.500	.375	.125
16	5	3	-	N	N	O	O	O	O	N	O	.625	.375	-
14	3	3	2	T	N	T	O	O	O	N	N	.375	.375	.250
7	3	5	-	N	N	N	N	O	O	N	O	.375	.625	-
15	4	3	1	N	N	O	O	T	O	O	N	.500	.375	.125
12	7	-	1	O	T	O	O	O	O	O	O	.875	-	.125
11	2	5	1	N	N	O	N	O	T	N	N	.250	.625	.125
10	6	-	1	O	O	O	N	O	T	O	O	.750	.125	.125
9	7	-	1	O	T	O	O	O	O	O	O	.875	-	.125
8	7	1	-	O	O	N	O	O	O	O	O	.875	.125	-
5	6	2	-	O	O	O	O	O	O	N	N	.750	.250	-
17	87	37	12									.639	.272	.088

Proportion of O, N, T Choices on Each Target Expression

p(X)	1	2	3	4	5	6	7	8	Σ
O	O=9 p(O)=.529	6 .352	13 .764	12 .705	14 .823	13 .764	9 .529	11 .647	87 .639
N	N=6 p(N)=.352	8 .470	3 .176	5 .294	1 .058	2 .117	7 .411	5 .294	37 .272
T	T=2 p(T)=.117	3 .176	1 .058	- -	2 .117	2 .117	1 .058	1 .058	12 .088

136

Decoding/Dissimilarity Judgments:

Own Sets

Again, the chance proportion of optimal success is .19. The observed range of O response for individual Ss is .500 (4) to 1.00 (16), (Table 3). It is noteworthy that, of the 23 N selections (summing across Ss and message sets), 12 were consistent with Ss' encoding. The proportion of O sets for any given encoded target was .357 (5/14) to .857 (12/14)--see Column marginals, Table 3. The rank of targets according to proportion of O decoding is not the same as that for encoding. In the decoding task, the targets order from most to least O:

3, 4, 5 > 6 > 8 > 7, 1 > 2

whereas for encoding

5 > 3, 6 > 4 > 8 > 7, 1 > 2

All difference of rank is dependent upon targets 3, 4, 5 and 6. One aspect resolving the discrepancy is exemplified by decoding task #53 compared to encoding target 6. In the latter case, Ss select kinematic expressions as being most opposite to the target. In decoding, Ss select the unlabeled expression as being "odd".

Predicted

Optimality of predicted sets varies greatly by individuals and by subcategories. The range of optimality for individuals is illustrated by .436 (24/55) for S 18 and .792 (42/53) for S 16 (Table 4).

E-defined

Optimality regarding E-defined sets is simply the sum of own and predicted optimality ratios by individuals. It is instructive to look at the proportion of O decodings by subcategories of message sets.

Faces: As described earlier, the message sets which fall in this category reflect four expressions from one face of the formal model. The expressions might be all first derivatives, or second derivatives, with respect to time, kinematic or dynamic, integrated with respect to distance or not. The model would predict that, given four expressions from one face, Ss would have no basis for making oddity judgments. In inspecting the data of all Ss for given decoding sets in the faces category, no a posteriori inferences could be drawn.

Table 3

Proportion of Optimal, Non-optimal, and Toss-up Choices on "Own" Sets, Decoding Task

Sum of Choices Per Subject				Target Expressions										
S	O	N	T	1	2	3	4	5	6	7	8	p(O)	p(N)	p(T)
1	6	1	1	O	O	O	O	T	N	O	O	.750	.125	.125
3	4	2	2	O	N	O	N	O	T	T	O	.500	.250	.250
2	7	1	-	O	N	O	O	O	O	O	O	.875	.125	-
4	4	2	2	N	T	O	O	O	O	N	T	.500	.250	.250
8	7	1	-	O	O	N	O	O	O	O	O	.875	.125	-
10	6	1	1	O	O	O	N	O	T	O	O	.750	.125	.125
12	7	-	1	O	T	O	O	O	O	O	O	.875	-	.125
16	5	3	-	N	N	O	O	O	O	N	O	.625	.375	-
14	3	3	2	T	N	T	O	O	O	N	N	.375	.375	.250
13	8	-	-	O	O	O	O	O	O	O	O	1.000	-	-
17	4	4	-	N	N	O	O	O	O	N	N	.500	.500	-
15	4	3	1	N	N	O	O	T	O	O	N	.500	.375	.125
5	6	2	-	O	O	O	O	O	O	N	N	.750	.250	-
	78	23	11									.696	.205	.098

Proportion of O, N, T Choices on Each Target Expression

		1	2	3	4	5	6	7	8		
O		9	5	12	12	12	11	8	9	f	78
		.643	.357	.857	.857	.857	.786	.571	.643	p	.696
N		4	6	1	2	-	1	5	4	f	23
		.286	.428	.071	.143	-	.071	.357	.286	p	.205
T		1	3	1	-	2	2	1	1	f	11
		.071	.214	.071	-	.143	.143	.071	.071	p	.098

Table 4

Number of Optimal Choices per Number of Oddity (Decoding) Choices, Predicted Sets

S	E-defined	Own	Predicted	a	c	Category (b_1)	(b_2)	(b_3)	(d_1)	(d_2)	(d_3)
1	47/62	8/8	39/54	2/6	6/8	6/6	6/7	5/6	6/8	6/8	4/5
2	48/62	7/8	41/54	2/6	5/7	5/5	6/8	6/6	6/8	5/7	6/7
3	36/62	5/7	31/55	3/6	4/8	3/6	5/7	1/4	4/8	7/8	4/8
4	36/62	6/8	30/54	2/6	2/8	0/6	2/4	8/8	4/8	6/7	6/7
5	46/62	6/6	40/56	2/6	6/7	5/5	8/8	4/8	5/8	4/7	6/7
8	35/57	5/8	30/49	3/6	2/7	4/7	4/7	6/6	6/6	3/6	2/4
9	37/62	7/8	30/54	0/6	3/8	6/8	5/7	1/1	5/8	6/8	6/8
10	46/62	6/8	40/54	2/6	5/8	1/1	3/7	8/8	8/8	6/8	7/8
12	48/62	8/8	40/54	2/6	3/8	0/0	8/8	8/8	6/8	7/8	6/8
13	50/61	8/8	42/53	3/6	7/7	6/8	6/8	0/0	7/8	7/8	6/8
14	34/62	4/7	30/55	1/6	3/7	4/7	4/8	5/7	4/8	5/5	4/7
15	26/62	2/7	24/55	2/6	2/7	4/6	4/8	4/8	2/6	5/7	2/7
16	30/62	5/8	25/54	2/6	6/8	3/6	2/6	0/4	5/8	5/8	2/8
17	37/62	4/5	33/57	1/6	3/8	3/7	7/8	4/5	5/8	6/7	4/8
	556/862	81/104	475/758	.321	.537	.641	.693	.750	.829	.764	.660
	.645	.778	.626				.697			.748	
		(.285-1.00)	(.436-.792)								

(i.e., no hypotheses) about the nature of response strategies. What is most important is that, within this category, the correspondence between O and given decodings is of the magnitude of .321 (27/84) when Toss-ups are considered as O. If toss-ups are not considered O, the proportion reduces to .142 (12/84). In addition, it may be noted that never did a S encode a message set consistent with a face of the model.

The second category of decoding message sets also affords no bases for predictions from the model. (Note: it is true that predictions could be made that "odd" expressions would be selected on the basis of labelability). The proportion of O decodings are much higher than for the face category, however; .549 (61/111).

The apparent bases for judgments depended upon the construction of the message sets. Apparent distinctive features were integration with respect to distance, whether expressions were first or second derivatives with respect to time and, to a lesser extent, labelability of expressions. The kinematic-dynamic distinction was not represented.

One aspect of the decoding oepration for these sets is not reflected in the few statistics above but is most readily observed in Table 5. Response to each of the 8 stimulus sets tended to be polarized between two alternatives within each set. (Only 11 of the 112 Rs were not so distributed). Again, if predictions were to made from distances in the model, one would expect chance distribution of Rs across all expressions. Since only 2 of 4 expressions were selected as Rs in each set, no statistics are required to demonstrate that Ss can make discriminations in some fashion not predicted by the model, but rather consistent with dissimilarity judgments.

A third category was predicted to reflect Ss' dissimilarity judgments. In this set, the model clearly predicts which expressions within each set should be selected as "odd" on the basis of distances between vertices in the model. Congruence with the hypothesis is reflected in the fact that of the 334 Rs to the 24 sets, only 60 did not conform exactly to the predictions (274 did conform); that amounts to an average of 2.5/14 not correctly predicted per message set. By comparison, the proportion of O judgments based on dissimilarity judgments is (246/334). It should be further noted that distances in the formal model (and in the psychological model since that equates with the formal) inevitably overrides labelability of expressions.

Table 5

Number of Optimal choices/oddity selections, decoding task, all E-defined message sets.

Categories	Ss 1	2	3	4	5	8	9	10	12	13	14	15	16	17		
(a) faces	2/6	2/6	3/6	2/6	2/6	3/6	0/6	2/6	2/6	3/6	1/6	2/6	2/6	1/6	27/84 .321	
(c) 3E-3D	6/8	6/8	4/8	2/8	7/8	3/8	3/8	5/8	3/8	7/7	4/8	2/8	6/8	3/8	61/111 .549	
(b_1) 3E,2D,1S	8/8	7/8	4/8	1/8	8/8	5/8	6/8	6/8	8/8	6/8	5/8	4/8	5/8	4/8	79/112 .705	
(b_2) 3E,2D,1S	7/8	6/8	5/8	5/8	8/8	4/7	5/8	4/8	8/8	6/8	4/8	4/8	2/8	7/8	76/111 .682	.736
(b_3) 3E,2D,1S	7/8	8/8	5/8	8/8	4/8	7/7	8/8	8/8	8/8	8/8	6/8	4/8	3/8	7/8	91/111 .819	
(d_1) 3D,2E,1S	6/8	6/8	4/8	4/8	5/8	6/8	5/8	8/8	6/8	7/8	4/8	3/8	5/8	5/8	74/112 .660	
(d_2) 3D,sE,1S	6/8	6/8	7/8	7/8	5/8	3/6	6/8	6/8	7/8	7/8	6/8	5/8	5/8	6/8	82/110 .745	.687
(d_3) 3D,2E,1S	7/8	7/8	4/8	7/8	7/8	4/7	6/8	7/8	6/8	6/8	4/8	2/8	2/8	4/8	73/111 .657	
	.758	.774	.580	.580	.741	.614	.596	.741	.774	.819	.548	.419	.483	.596		

A third category was predicted to reflect Ss' dissimilarity judgments. In this set, the model clearly predicts which expressions within each set should be selected as "odd" on the basis of distances between vertices in the model. Congruence with the hypothesis is reflected in the fact that of the 334 Rs to the 24 sets, only 60 did not conform exactly to the predictions (274 did conform); that amounts to an average of 2.5/14 not correctly predicted per message set. By comparison, the proportion of O judgments based on dissimilarity judgments is (246/334). It should be further noted that distances in the formal model (and in the psychological model since that equates with the formal) inevitably overrides labelability of expressions.

The remaining category affords the opportunity to force Ss into making discriminations according to the three dimensions of the model. In this instance, however, the model at least predicts the alternatives (at most, two). The proportion of O choices in this category is .687 (229/333). Predictions from the model would be that "odd" choices would be made consistent with the expressions which differ by two transformations (i.e., are opposed along the grand diagonal of the model). Of the 333 Rs to the 24 message sets, 24 were not consistent with the polarity hypothesis predicted by the model--the proportion consistent is therefore .927 (309/333). Typically, S opt for the same (common) choice of "odd" expression; 19 of the 24 sets may be so characterized. Of the aforementioned 19, 11 were consistent with oddity judged according to level of differentiation with respect to time, 7 with integration with respect to distance, 1 with kinematics-dynamics (partial integration of mass with respect to distance).

The general conclusion reached at this stage is that the communications task as a converging operation tends to reflect judgments by Ss consistent with (a) the Ss' dissimilarity judgments and (b) the logical model. Further, Ss tend to be consistent across the three tasks: encoding, decoding, dissimilarity judgments. The fact that Ss produce O choices with high frequency on the predicted tasks militates against a possible conclusion that Ss remember their encoding choices. In the decoding task, the dominant distinctive feature tends to be level of differentiation with respect to time. It further appears that physics "meaning" as exemplified by the kinematic-dynamic dimension as well as labelability are, for the decoding task compared to dissimilarity judgments, less predictive of optimality than is the formal model.

References:

Isaac, P. D. Dissimilarity judgments and multidimensional scaling configuration as indices of perceptual structure; a study of intra-individual consistencies. Publication of the Department of Psychology, the University of Michigan, Ann Arbor, Michigan, 68-3.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1966.

Rosenberg, S. & Cohen, B. D. Referential processes of speakers and listeners. Psychological Review, 1966, 73, 208-231.

Runkel, P. J. Cognitive similarity in facilitating communication. Sociometry, 1956, 19, 178-191.

Chapter 5

Perceptual Structure as a Function of Subject Matter Mastery

Findings described thus far indicate that Ss who have achieved considerable mastery in the subject matter of mechanics (i.e., graduate students in physics) respond to mathematical expressions in a manner predicted from the model used to formally generate these expressions. It is of interest to ask in what sense patterns of response in a judgment task reflect level of subject matter mastery.

To investigate this problem, 24 undergraduate students in physics and 24 undergraduate students in psychology were run in the second prototype experiment reported in chapter 3. Design, randomization and stimuli were the same as in that experiment.

Results and Discussion

Table 1 contains the intertrial correlations for both responses and response latency for physics and psychology undergraduates. The correlations for response data in the two groups of Ss are remarkably high and virtually indistinguishable from one another. In the case of latency the correlations are lower and less similar between the two groups. As in our earlier experiment, subsequent analyses were performed on trial 3 data.

Mean dissimilarity and standard deviations for each stimulus pair are presented in Table 2 for the physics undergraduates and Table 3 for the psychology undergraduates. Inspection of the data in these two tables reveals no uniform difference between the two groups of Ss.

The mean judgments for each group were then subjected to multidimensional scaling procedures. As might be expected from viewing the data in Tables 2 and 3, the configurations were almost identical in each case, and moreover, they were quite like similar configurations for the graduate student data presented in chapter 3. The configuration for the mxt stimulus set for the physics undergraduates appears in Fig. 1 and the comparable configuration for the psychology undergraduates appears in Fig. 2.

The results of the judgment task appear rather startling on first glance. Evidently, undergraduate students in a psychology course (sophmores by and large) can make judgments of similarity among complex alphanumeric expressions in physics and mathematics even though

Table 1

Intertrial Correlations for Group Ratings
Physics and Psychology Undergraduates

Stimulus Set	Physics			Psychology		
	Trials			Trials		
	1-2	1-3	2-3	1-2	1-3	2-3
αβγ	.97	.97	.98	.96	.97	.98
mxt	.94	.92	.98	.95	.92	.97
gm-cm	.94	.95	.97	.93	.92	.96
color	.98	.98	.99	.98	.98	.99

Intertrial Correlations for Response Latency
Physics and Psychology Undergraduates

Stimulus Set	Physics			Psychology		
	Trials			Trials		
	1-2	1-3	2-3	1-2	1-3	2-3
αβγ	.71	.70	.82	.49	.61	.65
mxt	.58	.60	.55	.58	.80	.70
gm-cm	.41	.39	.53	.55	.64	.52
color	.42	.57	.58	.26	.17	.58

Table 2

Mean Dissimilarity Rating and Standard Deviation
for Each Stimulus Pair on Last Trial
Physics Undergraduates

Category	Stimulus Pair	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
x	12	2.50	1.35	3.16	1.97
	34	2.66	1.37	3.29	1.94
	56	3.29	2.57	3.62	2.18
	78	3.04	2.59	3.50	2.08
y	51	5.70	2.29	6.04	2.17
	62	5.75	2.30	6.08	2.20
	73	5.33	2.31	5.79	2.60
	84	5.66	2.35	5.54	2.39
z	14	5.50	2.35	4.00	1.93
	23	5.41	2.24	5.33	2.01
	58	5.83	2.58	5.50	1.93
	67	5.25	2.26	5.79	2.16
xy	16	6.50	1.86	7.54	1.84
	25	6.41	2.01	7.04	1.80
	38	6.08	2.20	6.50	2.00
	47	6.75	2.00	6.50	2.24
xz	31	6.54	1.99	6.66	1.92
	42	6.66	2.05	7.04	1.70
	75	6.79	2.30	7.58	1.95
	86	6.25	1.84	7.00	1.95
yz	18	5.79	3.36	6.08	3.54
	27	5.62	3.13	6.33	3.04
	36	8.33	1.94	7.33	2.09
	45	7.95	2.29	7.54	2.18
xyz	53	8.16	2.25	8.33	1.99
	64	8.62	2.08	8.25	1.98
	71	6.37	3.37	7.75	2.89
	82	6.12	2.96	7.04	2.75

Table 2

(cont'd)

Category	Stimulus Pair	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
x	12	3.75	1.79	3.25	1.39
	34	3.66	1.78	3.50	1.47
	56	4.41	2.37	3.79	1.88
	78	3.70	1.87	3.33	1.43
y	51	4.04	2.11	3.70	1.57
	62	4.50	2.16	3.66	1.94
	73	3.91	1.71	3.29	1.42
	84	4.12	2.23	3.58	1.47
z	14	4.37	2.01	5.04	1.80
	23	3.87	1.54	4.62	2.01
	58	4.50	2.16	4.58	1.95
	67	4.70	2.01	5.20	2.08
xy	16	6.95	2.17	5.45	2.06
	25	7.04	2.03	5.58	2.01
	38	6.70	1.98	5.29	2.09
	47	6.58	2.50	5.75	1.98
xz	31	6.66	1.85	6.54	2.02
	42	7.08	2.12	6.45	2.08
	75	6.41	1.99	6.29	1.94
	86	7.41	2.20	6.33	1.92
yz	18	4.45	2.02	6.87	2.21
	27	5.33	1.88	6.66	2.46
	36	6.29	2.08	7.12	2.32
	45	6.45	2.30	6.70	2.49
xyz	53	8.25	1.79	9.04	2.27
	64	7.75	2.15	9.08	2.20
	71	7.70	2.21	9.29	1.96
	82	7.04	2.45	9.08	2.32

Table 3

Mean Dissimilarity Rating and Standard Deviation for Each Stimulus Pair or Last Trial Psychology Undergraduates

Category	Stimulus Pair	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
x	12	2.41	.97	2.73	1.35
	34	3.39	2.33	3.12	1.84
	56	2.70	1.68	3.58	1.81
	78	3.50	2.82	3.60	2.21
y	51	5.73	2.43	5.16	2.18
	62	6.08	2.33	4.83	1.92
	73	5.86	2.56	5.75	1.82
	84	5.62	2.20	5.66	2.38
z	14	4.39	1.72	4.52	2.29
	23	4.45	1.69	4.04	2.01
	58	4.56	1.80	4.83	2.09
	67	5.00	1.90	4.37	1.68
xy	16	7.91	1.81	7.26	1.95
	25	6.04	2.36	5.91	1.47
	38	6.20	2.63	7.08	2.21
	47	7.20	2.41	6.71	1.76
xz	31	5.65	2.10	6.68	2.21
	42	5.00	2.64	6.29	2.25
	75	5.30	2.43	6.43	2.42
	86	4.54	1.84	6.26	2.89
yz	18	9.34	1.64	8.30	2.05
	27	8.08	2.29	8.33	1.90
	36	6.50	2.82	6.58	1.93
	45	6.87	2.41	7.30	2.03
xyz	53	7.25	2.64	7.73	1.30
	64	7.52	2.27	7.95	2.06
	71	9.66	1.88	9.83	1.81
	82	8.54	2.22	8.75	1.80

Table 3

(cont'd)

Category	Stimulus Pair	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
x	12	4.83	2.35	3.08	1.34
	34	5.16	2.59	2.65	1.19
	56	5.26	2.57	3.04	1.55
	78	4.91	2.48	3.33	1.68
y	51	3.25	2.62	3.29	2.25
	62	3.60	2.70	2.66	1.30
	73	3.30	2.28	2.69	1.79
	84	3.13	2.32	2.58	1.17
z	14	2.75	1.77	5.50	2.53
	23	3.20	2.01	5.45	2.84
	58	3.26	2.24	5.30	2.63
	67	3.26	2.04	5.86	2.85
xy	16	7.12	2.62	5.16	2.05
	25	6.81	2.57	4.66	1.63
	38	7.08	2.87	5.45	2.26
	47	7.04	2.18	5.62	2.61
xz	31	6.83	2.23	7.33	2.16
	42	6.86	2.61	6.83	2.46
	75	7.34	2.34	7.12	2.52
	86	6.87	2.77	7.54	2.37
yz	18	3.86	2.47	7.[illegible]6	2.56
	27	4.86	2.96	7.79	2.84
	36	5.63	2.87	7.90	2.44
	45	4.43	3.07	7.83	2.66
xyz	53	7.00	2.52	9.79	2.08
	64	6.71	2.74	9.45	2.50
	71	7.21	3.02	9.69	2.03
	82	7.39	2.75	9.95	1.82

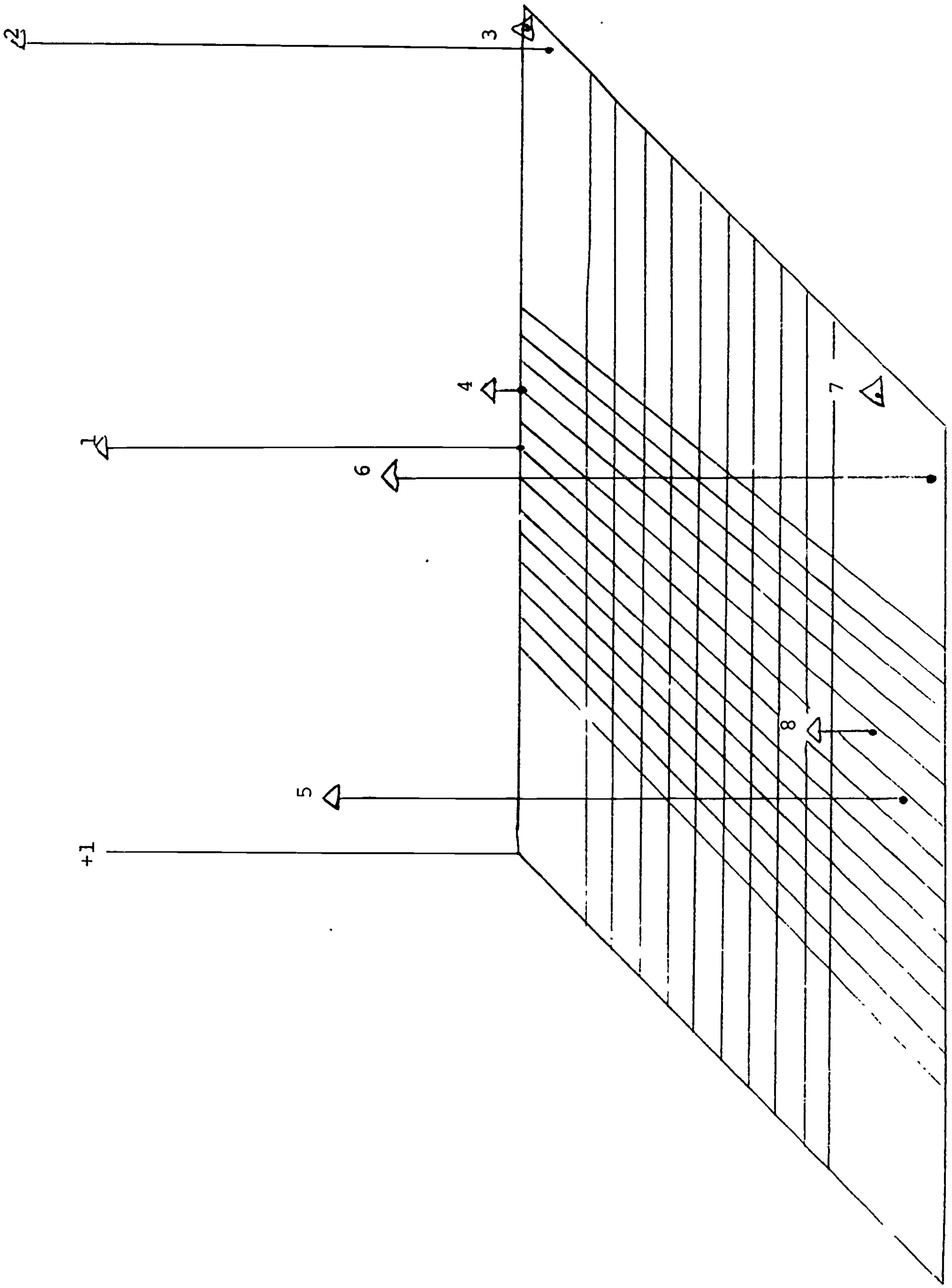


Fig. 1. Metric Scaling of MXT Dissimilarity Judgment - Physics Undergraduates

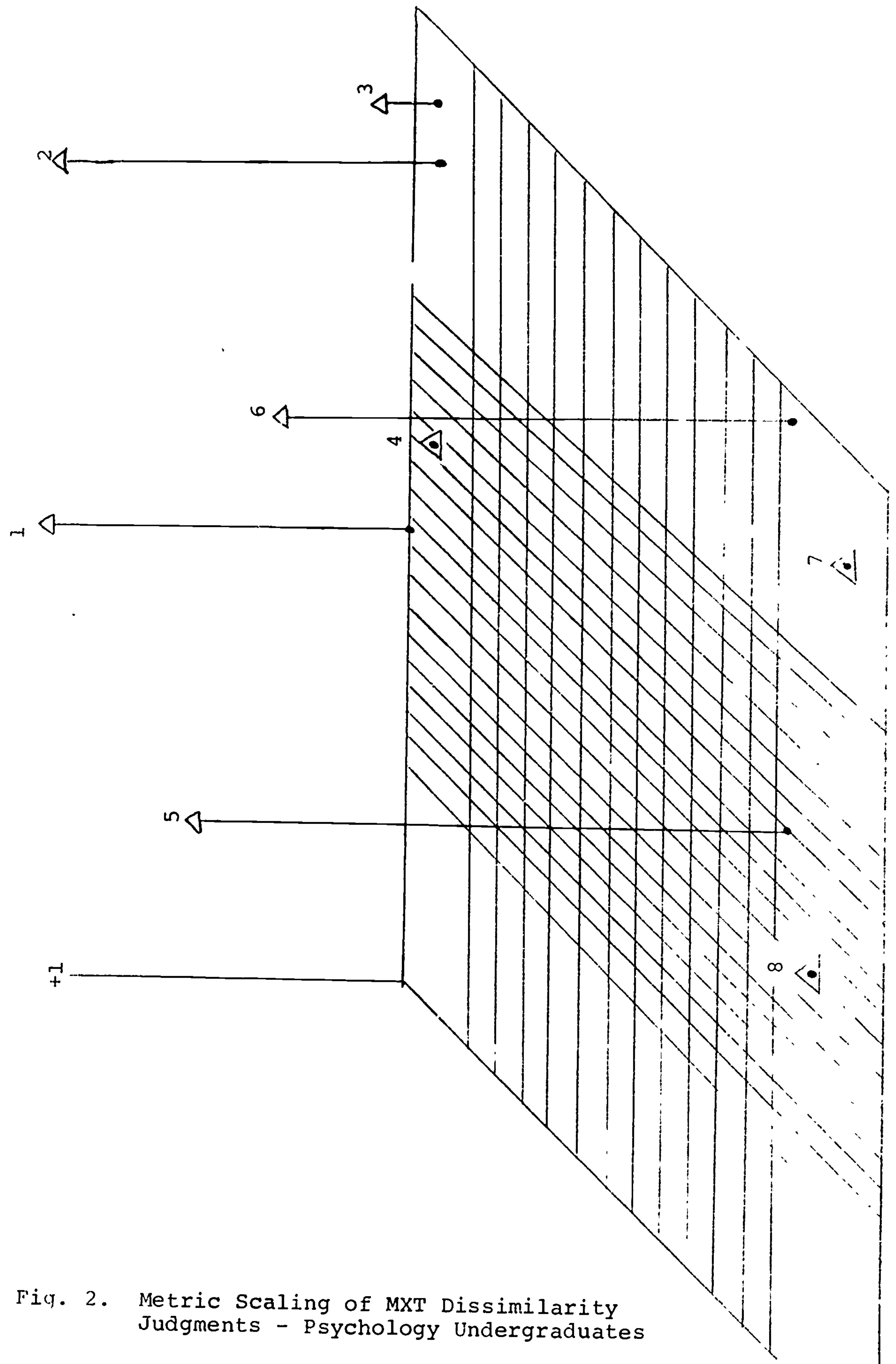


Fig. 2. Metric Scaling of MXT Dissimilarity Judgments - Psychology Undergraduates

they have little if any knowledge of the meaning of these expressions. Of course, most college sophomores have probably had a physics course somewhere in their background and maybe even a calculus course (most certainly an algebra course), but it would seem rather unparsimonious to suppose that the psychologist knows physics (or that the physicist does not). Rather, what is probably true is that psychology as well as physics undergraduates perform the judgment task on the basis of some set of distinctive features. Since the response patterns of the two groups are remarkably alike we can assume Ss detect much the same stimulus features regardless of background experience. This does not mean, however, that the knowledge of the physics student is the same as the knowledge of the psychology student with regard to the stimulus expressions. Rather, it means that what the physics student did, produced the same kind of response pattern that appeared in the data from the psychologists. It also means that the task of judging the similarity among abstract expressions as reflected in the group data presented in Table 1 does not serve to detect the presence or absence of previous educational experience.

Because of the comparability of data in the two groups of Ss we cannot assume that all physics undergraduates performed the task by utilizing their knowledge of physics. The performance of the psychology students leaves open the probability that physics students may also have chosen the simpler strategy of responding to the presence and absence of major distinctive features (e.g., $\int$, dt, gm).

The latency data were examined to see whether they might shed light on the way each group of Ss performed the judgment task. Mean latency and standard deviations for each stimulus pair are given in Tables 5 and 6 for the two groups of Ss. Inspection of the data in these tables indicates that psychology students spent about the same amount of time per stimulus pair on each of the four tasks, suggesting that in each case the strategy they employed was to count the number of differences (based upon a distinctive feature analysis) between the two stimuli comprising a pair and then assign this difference a number. The physics students, however, spent more time on the three alpha-numeric stimulus sets than they did on the color set. Moreover, the time the physics students spent on the color set was quite similar to the time the psychology students spent on this same set.

Table 5

Mean Response Latency and Standard Deviation
For Each Stimulus Pair on Last Trial
Physics Undergraduates

Category	Stimulus Pair	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
x	12	2.04	.71	2.30	1.02
	34	2.59	1.10	3.07	1.58
	56	2.97	1.14	3.30	1.76
	78	3.46	1.46	3.60	1.56
y	51	2.99	.98	4.02	3.18
	62	3.66	1.63	3.45	1.92
	73	3.34	1.23	3.63	1.79
	84	2.90	1.07	3.29	1.62
z	14	2.49	.90	3.11	1.64
	23	3.08	1.20	3.43	1.55
	58	3.37	1.41	3.78	2.19
	67	3.81	2.00	3.83	2.35
xy	16	3.55	1.48	3.32	1.56
	25	3.38	1.36	3.44	1.38
	38	3.72	1.65	3.76	1.57
	47	3.59	1.78	3.71	1.91
xz	31	3.25	1.34	3.29	1.75
	42	3.00	1.20	3.55	2.10
	75	3.38	1.26	4.31	2.55
	86	4.05	2.07	4.50	2.34
yz	18	3.26	1.37	4.09	2.97
	27	3.92	1.66	4.84	4.44
	36	3.82	1.74	3.62	2.32
	45	3.32	2.02	3.81	2.26
xyz	53	3.44	1.58	3.53	1.74
	64	3.78	1.91	4.24	4.64
	71	3.27	1.52	3.95	2.25
	82	3.49	1.67	3.66	1.94

Table 5

(cont'd)

Category	Stimulus Pair	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
x	12	2.84	1.73	2.12	1.38
	34	3.65	1.78	2.10	1.34
	56	3.51	1.79	2.14	.98
	78	3.19	1.56	2.19	1.70
y	51	2.83	1.27	2.13	1.08
	62	3.79	2.37	2.29	1.44
	73	3.58	1.29	2.19	1.43
	84	2.97	1.88	1.89	1.11
z	14	2.81	1.42	2.57	1.50
	23	3.45	1.84	2.24	1.00
	58	3.25	1.66	2.50	1.98
	67	3.98	1.95	2.39	1.49
xy	16	3.76	2.21	2.52	1.45
	25	3.57	1.86	2.49	1.38
	38	3.85	2.34	2.66	1.78
	47	3.90	2.13	2.54	1.03
xz	31	3.59	1.91	2.76	1.90
	42	4.80	6.89	2.44	1.60
	75	3.71	2.11	2.55	1.71
	86	3.87	2.14	2.35	.94
yz	18	2.84	1.39	2.41	1.14
	27	3.95	2.10	2.46	1.29
	36	3.55	1.80	2.60	2.01
	45	3.49	1.69	2.48	1.72
xyz	53	3.89	1.[illegible]1	2.05	.80
	64	3.81	2.37	2.28	1.63
	71	3.42	1.47	2.21	1.46
	82	3.73	2.06	2.00	1.00

Table 6

Mean Response Latency and Standard Deviation
for Each Stimulus Pair on Last Trial
Psychology Undergraduates

Category	Stimulus Pair	αβγ $\bar{x}$	αβγ s	mxt $\bar{x}$	mxt s
x	12	2.34	.88	2.33	.53
	34	2.99	.95	3.14	.89
	56	2.79	1.02	2.86	.75
	78	2.96	.80	3.13	.91
y	51	2.59	.76	2.61	.62
	62	2.78	.86	2.77	.91
	73	2.82	.82	2.89	.86
	84	2.84	.97	3.02	.91
z	14	2.69	.90	2.70	.93
	23	2.74	.79	3.05	.86
	58	2.84	.96	3.15	.78
	67	2.89	.83	3.00	.85
xy	16	2.79	.98	2.87	.75
	25	2.86	.83	2.90	.92
	38	2.90	.96	3.08	1.06
	47	2.87	.98	3.07	.96
xz	31	2.59	.88	2.72	.90
	42	2.70	.88	3.04	.75
	75	3.00	1.19	3.01	.78
	86	2.90	.93	3.19	.86
yz	18	2.53	.80	2.71	.74
	27	2.60	1.08	3.10	1.06
	36	2.99	1.02	3.26	.93
	45	2.83	1.07	2.92	.80
xyz	53	3.02	.92	3.07	.84
	64	2.67	.93	2.91	.97
	71	2.54	.92	2.35	.84
	82	2.93	.90	2.76	.75

Table 6

(cont'd)

Category	Stimulus Pair	gm-cm $\bar{x}$	gm-cm s	color $\bar{x}$	color s
x	12	2.50	.72	2.29	1.03
	34	2.61	.79	2.03	.73
	56	2.66	.70	2.12	.93
	78	2.98	.87	1.83	.75
y	51	2.37	.68	2.20	.93
	62	2.46	.64	1.86	.61
	73	2.89	.69	1.97	.50
	84	2.36	.54	1.82	.76
z	14	2.33	.76	2.31	.81
	23	2.62	.84	2.33	.92
	58	2.46	.67	2.38	.95
	67	2.73	.73	2.26	.88
xy	16	2.72	.96	2.38	.88
	25	2.57	.87	2.19	.69
	38	2.77	.69	2.39	.99
	47	2.85	.83	2.22	.88
xz	31	2.59	.78	2.47	.87
	42	2.65	.73	2.26	.94
	75	2.88	1.02	2.39	.79
	86	2.60	.76	2.16	.83
yz	18	2.62	.71	2.37	.70
	27	2.67	.64	2.42	.85
	36	2.91	.87	2.21	.69
	45	2.68	.92	2.34	.95
xyz	53	2.46	.70	1.80	.52
	64	2.79	.85	2.11	.79
	71	2.54	.87	2.08	.81
	82	2.60	.78	2.04	.85

In order to examine the comparability of latencies between stimulus sets in each group, mean RL was computed on each trial across all stimulus pairs. The graphs of mean RL by trial for each group appear in Figs. 3 and 4. These figures reveal that RL is uniformly less for psychology students than for physics students and less for the color set than for the three stimulus sets involving the alpha-numeric symbol structure.

We conclude from an examination of group data (both response and response latency) that the knowledge required to judge the similarity among expressions in physics is consistent with the knowledge required to distinguish among rather arbitrary, but nevertheless regular, stimulus configurations. While having increased knowledge of the potential meaning of these expressions results in taking more time to perform the task, the end result is much the same whether one understands the technical vocabulary involved or not.

Individual Subject Analysis

In order to determine whether real differences between the two groups of subjects might be masked by group data, a separate analysis was performed on a sample of subjects from each group. This analysis was conducted by utilizing an individual difference multidimensional scaling program recently published by the Bell Telephone Laboratories (Carroll & Chang, 1971).

This program performs metric multidimensional scaling on individual subjects and provides both a group stimulus space (much in the fashion of the multidimensional scaling program utilized earlier) as well as an individual stimulus space for each subject, plus a multidimensional space in which subjects are plotted in relation to one another based on their responses to the stimulus set.

Examination of the scaling solutions for the six (out of 24) physics students sampled on the mxt and color tasks revealed that the three dimensional weights for the color task lie within a smaller area of space than for the mxt task. There was almost no overlapping for these 2 sets of weights. More explicitly, if the mean point for each of the tasks is computed, and the variance weighting for each task is computed as the average squared Euclidean distance of each subject's point weighting from this mean, then this variance for the color task is much less than for the mxt task. This means that the assignments of axis weighting are more homogeneous for the physics students in the color task than the mxt task. Consequently, we can assume that the physics students have structured the color task stimuli more

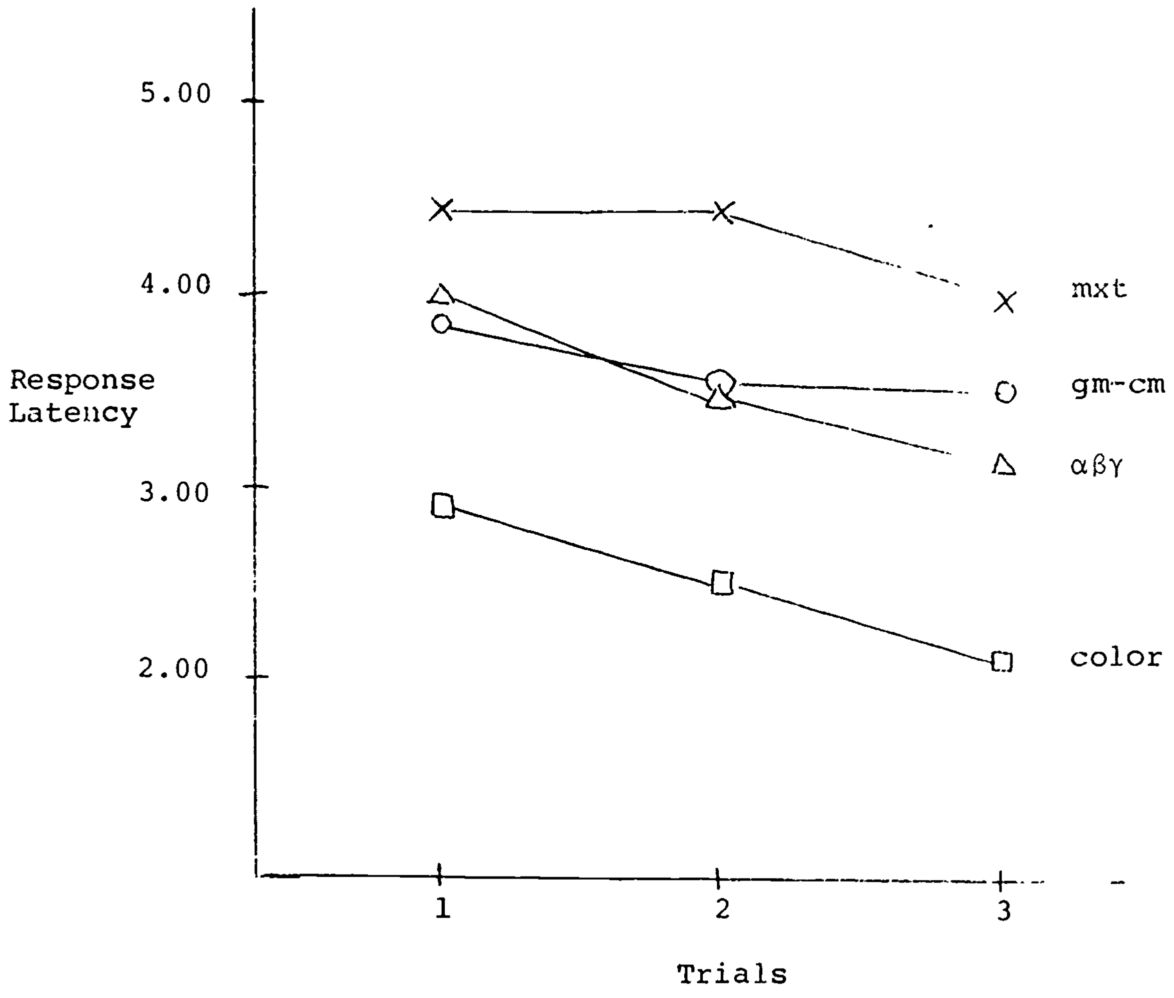


Fig. 3. Mean Latency (averaged over all stimulus pairs by trial)

Physics Undergraduates

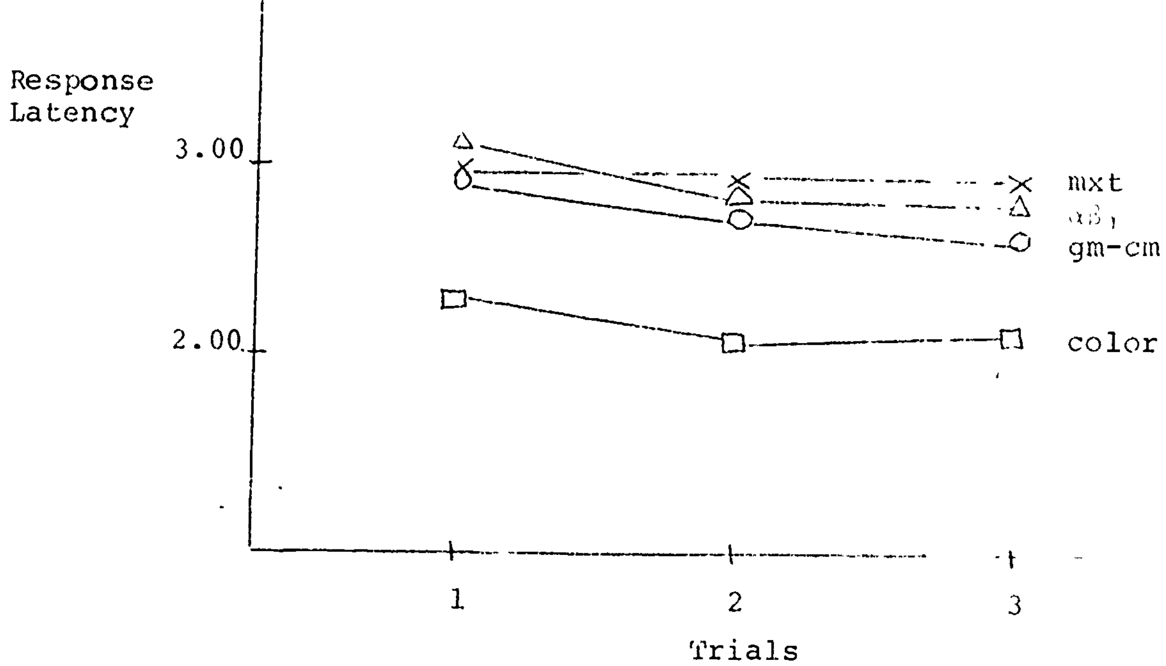
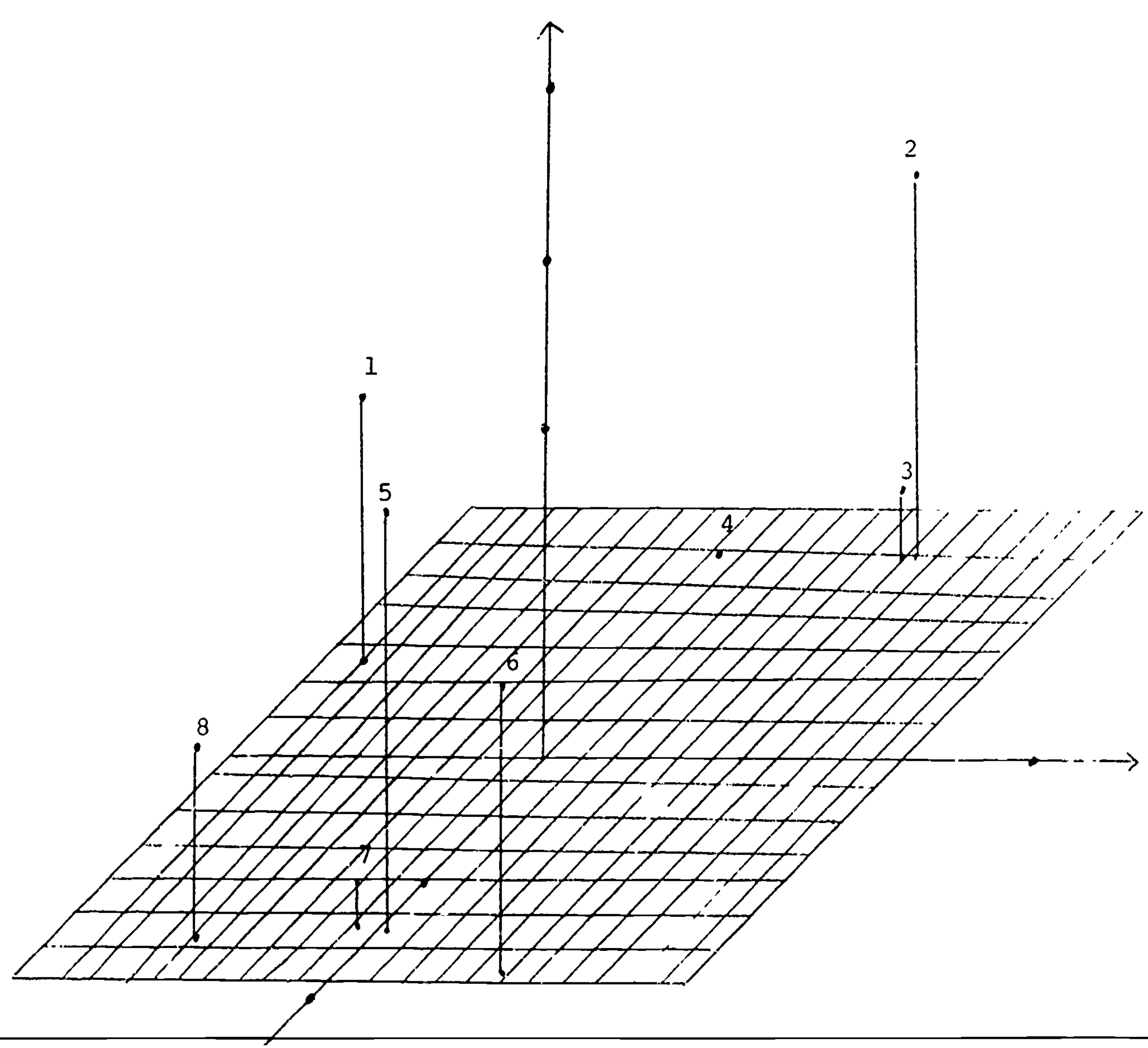


Fig. 4. Mean latency (averaged over all stimulus pairs by trial)

Psychology undergraduates



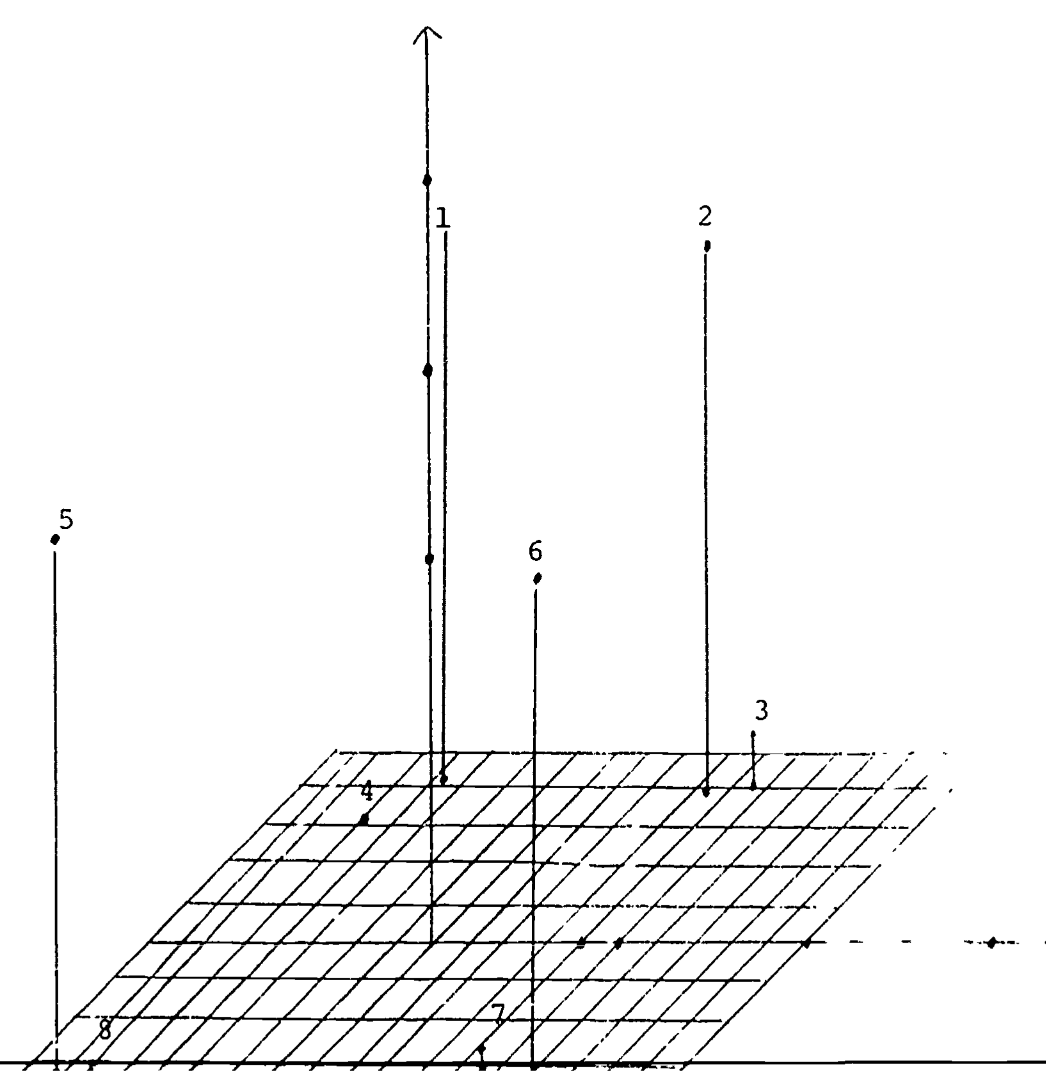
difference in the way the stimuli for the two tasks were structured in 3 dimensions.

The six physics and psychology students were then compared directly on the mxt task (see Fig. 5). The variance of the psychology students was much less than for the physics students. There was much more agreement regarding the structuring of the mxt stimuli for the psychology students. There was also little (possible one point) overlap in the assignment of relative weights to the mxt stimuli, for the two groups of subjects (the psychology students' points are closely clustered around the Z axis).

A comparison on the color task for the two groups of subjects reveals much intersecting of the two sets of weights (see Fig. 6). There was much agreement among the two groups of subjects regarding the structure of the stimuli in the color task. The group variance for the two sets of subjects is about the same, signifying that the range of variation in individual subject judgments in the two groups was similar. There was nothing about the color task that enabled the members of either group to weight the stimuli homogeneously within a selected area of three dimensional space.

The individual subject group stimulus spaces for three psychology subjects in the mxt task were examined. All of these subjects structured the stimuli in recognizable cubes, the vertices being positioned relative to one another consistently. What differed in this representation among the three subjects was the length of the edges (and hence diagonals and grand-diagonals). While all the subjects viewed the data as similarly structured, the relative interpoint similarity distances differed.

In the case of the physics students sampled from the mxt task, all the representations are also recognizable



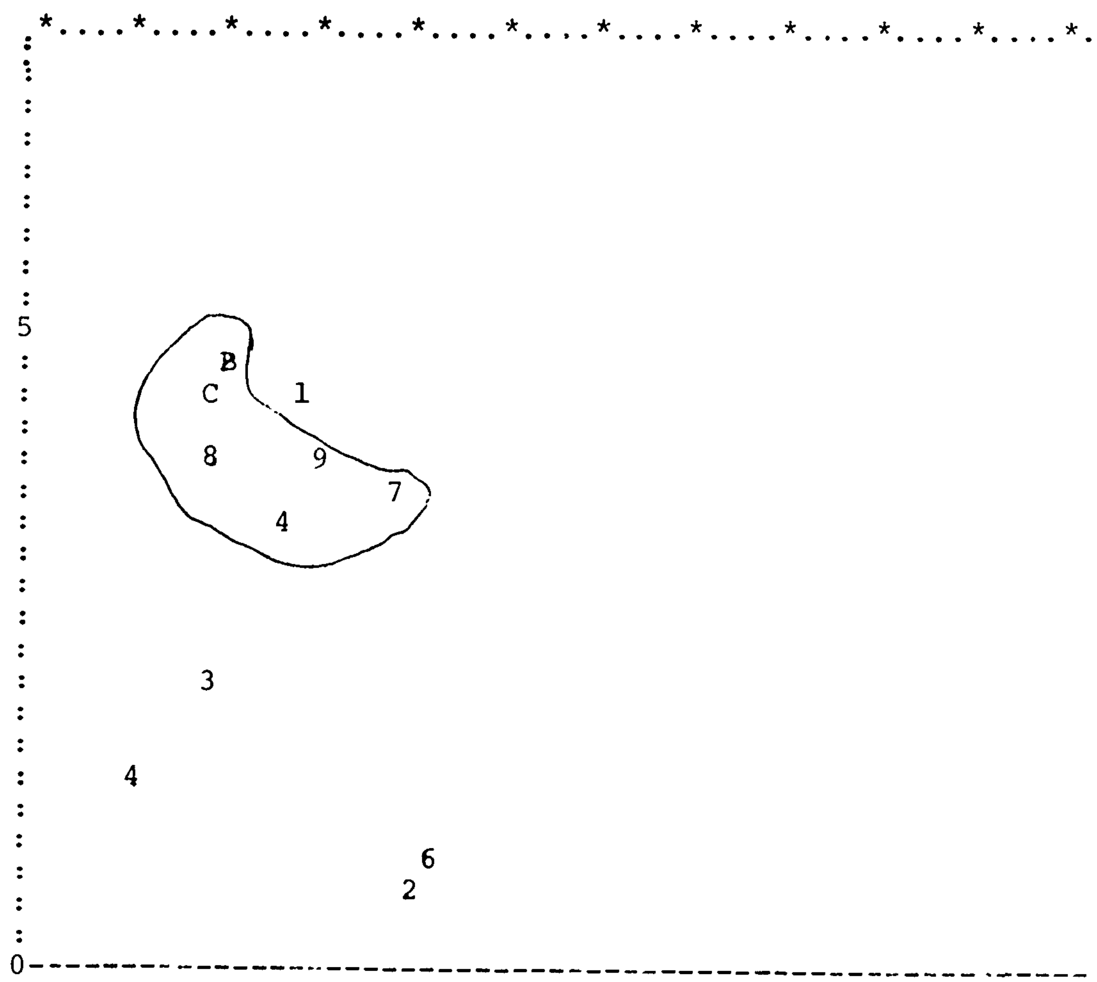


Fig. 5.
Normalized A-Matrix No. 1 -- Dimensions 1 and 3.

MXT - Physics (pts. 1-6) and Psychology (pts. A, B, C, 7, 8, 9) Undergraduates.

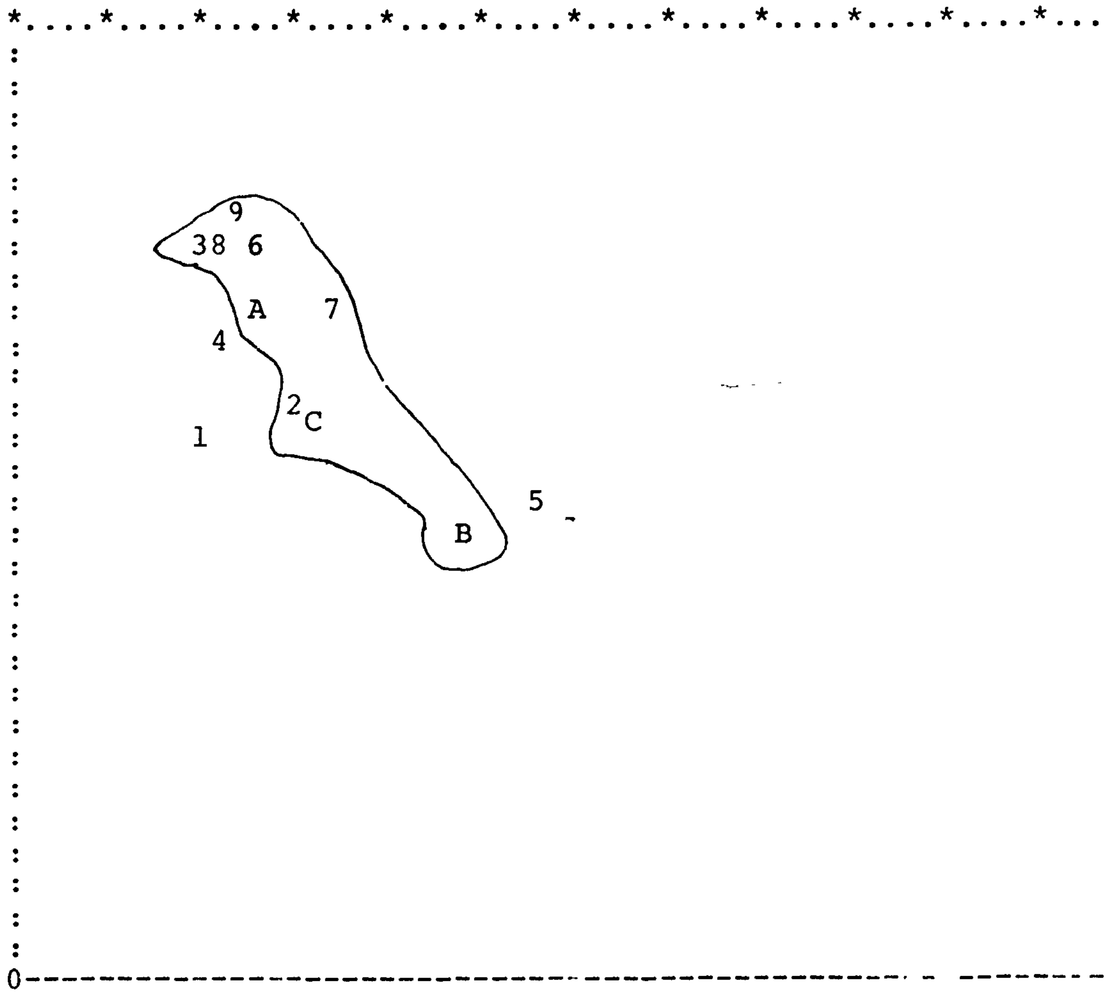


Fig. 6.

Normalized A-Matrix No. 1 -- Dimensions 1 and 3.

Color - Physics (pts. 1-6) and Psychology (pts. A, B, C, 7, 8, 9) Undergraduates.

as cubes (but this time much more distorted). In some cases, edges are differently oriented in three dimensional space. For example, in one subject's representation, vertex 5 lies to the right of vertex 8 (and above, as it should), and vertex 4 lies to the left (and below) vertex 1. In another subject's graph, vertex 5 lies to the left of vertex 8 and vertex 4 is to the right of vertex 1, while the other vertices remain in the same relative correspondence to one another. In still another graph, the same relations hold as in the first one described but vertex 6 is to the right of vertex 7. The characterization that uniformly applies to all the graphs is the relative height positioning of all the vertices. Thus, vertices 1, 2, 5, 6 are always correctly positioned above vertices 4, 3, 8, 7, respectively.

Summary Analysis for Abstract Stimulus Sets

Thus far we have examined the performance of physics graduate students, physics undergraduate students, and psychology undergraduate students in response to stimulus sets constructed to represent abstract concepts in physics. We have seen that in general, all three groups of subjects respond much alike, especially with regard to judgment data. The distinctive features built into each stimulus set are revealed in scaling analyses and these analyses conform to our expectation based on the abstract model used to generate the stimuli.

The two major ways we can distinguish among the groups of subjects are with regard to latency and with regard to the variability among individuals within the groups. For each stimulus set, Ss responses are a general increasing (monotonic) function of the number of transformations between members of a stimulus pair. (Stimulus pairs on edges involved one transformation, stimulus pairs on diagonals involved two transformations, stimulus pairs on grand diagonals involved three transformations).

The latency data do not give evidence of an increasing monotonic function relating psychological judgments and physical dimensions of the model. However, there is discrepancy between the three groups of subjects, particularly between the psychology undergraduates and the physics graduate students. Latency data from the three subject groups are summarized in Fig. 7 which illustrates that with increasing educational experience subjects take more time to make their judgments on those stimulus sets for which their experience has given them specialized knowledge.

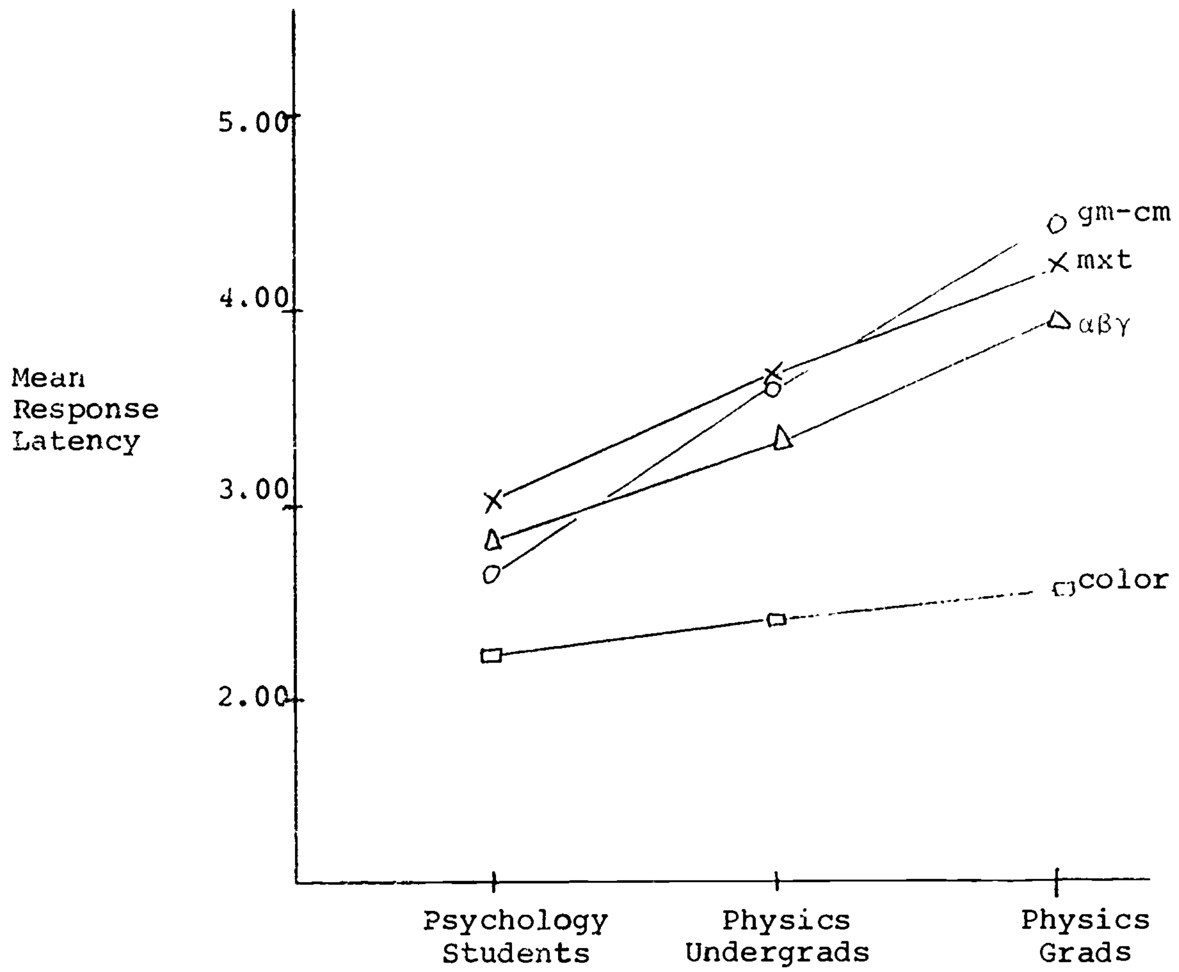


Fig. 7. Mean Latency Averaged Over All Stimulus Pairs on Last Trial.

Chapter 6

Perceptual Structure as a Function of Semantic Complexity

This study is analogous to experiments reported in chapters 3 and 4. It was composed of two distinct procedures -- one portion requiring the collection of similarity judgments on stimuli constructed from prototype problems in physics and the second portion being a communication task using these stimuli.

The stimuli in chapter 3 had little semantic content. They were constructed from syn...ctic features used to identify the general schematic form of concepts such as force, distance, and time. In this experiment, semantic content was introduced by having the concepts focus on prototype physical problems, in this case, the harmonic oscillator and the falling body.

The similarity judgment portion of the experiment was designed to check the validity of the model in Fig. 3, chapter 2, and to see whether the judgments are consistent with a feature analysis of physics "meaning" (whether physics students dichotomize according to kinematics and dynamics), and the labelability of expressions as physical concepts. The communication task, consisting of encoding and decoding operations, was designed to investigate the role of perceived similarity among prototype problem stimuli in constructing and decoding message sets containing those stimuli.

Fifteen graduate physics students were subjects in two experimental sessions -- the first session consisted of the similarity judgments and the encoding task; the second -- the decoding task. Dissimilarity judgments were obtained to each of three different stimulus sets listed in Table 1. Stimuli in the harmonic oscillator (H) and falling body (F) sets were obtained by means of procedures described in chapters 2 and 3. Briefly, if we represent the origin in Fig. 3 as x (the position of the mass, m) then in moving to any of the labeled vertices via a calculus operation we cannot follow a random path in the case of the prototype proble.. because in the general case, x may be a function of m and t, or both.

What controls the possible directions one can move from one vertex point to reach another is the "hierarchy of functional dependencies that is imposed by each prototype problem, and these must be worked out in each

case. The rule is that in tracing a path from the origin to a vertex one must never move along the axis of a variable x_n after having moved along the axis of a variable x_m when x_n is functionally dependent of x_m.

For these reasons the definitions of the vertices of the cube are correct only for the general form of expressions. For each prototype problem, we begin with known expressions of X and F and a known set of functional dependencies between variables.

Eight expressions were generated in the above fashion for each prototype problem (harmonic oscillator and falling body). Five of the eight expressions have convenient labels in physics. The remaining three expressions do not have physical labels and depend for their inclusion upon their logical (formal) generative relations. The five labeled expressions are:

	Falling Body	Harmonic Oscillator
Velocity	$v_o - gt$	$A\sqrt{\frac{k}{m}} \cos\sqrt{\frac{k}{m}}\, t$
Acceleration	$-g$	$-A(\frac{k}{m} \sin\sqrt{\frac{k}{m}}\, t$
Momentum	$-mgt$	$A\sqrt{mk} \cos\sqrt{\frac{k}{m}}\, t$
Force	$-mg$	$-kA \sin\sqrt{\frac{k}{m}}\, t$
Work	$-mgx_o - mgv_o t + \frac{1}{2}mg^2t^2$	$-\frac{1}{2}kA^2 \sin^2\sqrt{\frac{k}{m}}\, t$

The eight expressions for the two prototype problems are found in Table 1. The stimuli for the third set (labeled the combined set) consisted of the first four (labeled) stimuli in each of the two prototype problems. These expressions were chosen because they all lie on a single face in the model in Fig. 3 of chapter 2. Responses to pairwise combinations of these expressions thus permit a direct comparison of the two prototype problems.

The major questions of interest were (1) whether the structure of Ss responses to the more complex expressions would also reflect the structure of the model in Fig. 3 (chapter 2) and whether when faced with a mixed (or combined) response set, Ss would perceive structure among expressions based upon likeness of conceptual category (e.g., expressions for velocity belong together

Table 1

Experimental Stimuli for each Task

Position on Fig. 1 (Vertex)	Task: Harmonic Oscillator	Task: Falling Body	Task: Combined
1	$A\sqrt{\frac{k}{m}}\cos\sqrt{\frac{k}{m}}\,t$	$v_o - gt$	$v_o - gt$
2	$A\sqrt{mk}\cos\sqrt{\frac{k}{m}}\,t$	$-mgt$	$-mgt$
3	$-kA\sin\sqrt{\frac{k}{m}}\,t$	$-mg$	$-mg$
4	$-A(\frac{k}{m})\sin\sqrt{\frac{k}{m}}\,t$	$-g$	$-g$
5	$\frac{1}{4}A^2\sqrt{\frac{k}{m}}(\sqrt{\frac{k}{m}}t-\sin\sqrt{\frac{k}{m}}t\cos\sqrt{\frac{k}{m}}t)$	$-gx_o t-\frac{1}{2}gv_o t^2+\frac{1}{6}g^2t^3$	$A\sqrt{\frac{k}{m}}\cos\sqrt{\frac{k}{m}}\,t$
6	$-\frac{1}{4}A^2\sqrt{mk}(\sqrt{\frac{k}{m}}t-\sin\sqrt{\frac{k}{m}}t+\cos\sqrt{\frac{k}{m}}t)$	$-mgx_o t-\frac{1}{2}mgv_o t^2+\frac{1}{6}mg^2t^3$	$A\sqrt{mk}\cos\sqrt{\frac{k}{m}}\,t$
7	$-\frac{1}{2}kA^2\sin^2\sqrt{\frac{k}{m}}t$	$-mgx_o-mgv_o t+\frac{1}{2}mg^2t^2$	$-kA\sin\sqrt{\frac{k}{m}}\,t$
8	$-\frac{1}{2}A^2(\frac{k}{m})\sin^2\sqrt{\frac{k}{m}}t$	$-gx_o-gv_o t+\frac{1}{2}g^2t^2$	$-A(\frac{k}{m})\sin\sqrt{\frac{k}{m}}\,t$

regardless of the problem they are appropriate to) or likeness of problem category (e.g., the four concepts for the harmonic oscillator belong together as do the four concepts for the falling body problem.

In the judgment task, Ss were given five trials for one of the three stimulus sets. Stimuli were randomly ordered on each trial and were presented in pairs upon a frosted backlit screen at a 5 sec. interval.

In the encoding task, E presented S with the set of stimuli that S had previously scaled. E then selected one of these stimuli and declared it a "target." According to the target, S selected 3 of the remaining stimuli which made the target seem odd or out of context. Each of the 8 stimuli is selected, via a random order, as a target. Therefore, S produced 8 message sets.

All 15 Ss completed the decoding task. In this case, each S was given 32 message sets, each consisting of 4 of the stimuli from the task he had previously performed. S was told to choose the stimulus that seemed most odd of the four given, and to do this for each of the 32 message sets.

Results and Discussion

As in previous experiments, intertrial correlations were computed for both response and response latency. These data are presented in Table 2. An examination of Table 2 indicates that although correlations tend to be somewhat lo than in previous cases, correlations between adjacent trials are generally higher than between nonadjacent trials and the correlations between trials 4 and 5 and 3 and 4 are the highest in the table. As was found previously, the correlations for response latency are also generally lower and more erratic than correlations for responses.

Mean responses and the standard deviations for each stimulus pair were computed on the last trial of the task. These data are presented in Table 3 for all three stimulus sets. An examination of data in the table reveals that judgments seem to be a generally increasing function of stimulus transformations comprising each individual pair although there is considerably more variation within a given stimulus category than was found for abstract stimulus pairs. This is particularly true for the falling body prototype problem.

Table 2

Intertrial Correlations for Responses and Response Latencies

Trials	Responses H	Responses F	Responses C	Response Latencies H	Response Latencies F	Response Latencies C
1-2	.78	.88	.77	.16	.12	.51
1-3	.61	.89	.87	.29	.31	.28
1-4	.60	.85	.77	.31	.61	.41
1-5	.58	.82	.86	.30	.18	.49
2-3	.80	.97	.85	-.07	.38	.53
2-4	.84	.93	.88	.20	.45	.64
2-5	.79	.90	.84	-.15	.43	.46
3-4	.91	.95	.91	.31	.54	.50
3-5	.88	.94	.96	.20	.59	.37
4-5	.91	.94	.92	.03	.55	.46

H - Harmonic oscillator

F - Falling body

C - Combined

Table 3

Mean Responses and Standard Deviations for Each Stimulus Pair on Last Trial

Category	Stimulus Pair	Harmonic Oscillator		Falling Body		Combined	
		$\overline{X}$	s	$\overline{X}$	s	$\overline{X}$	s
x	12	4.40	3.20	6.25	2.87	1.20	0.45
	34	3.40	2.07	1.80	0.84	3.60	2.40
	56	5.00	3.39	1.60	0.55	6.25	2.87
	78	5.80	2.38	1.80	0.45	3.40	2.07
y	15	6.80	4.38	6.00	1.41	3.40	2.60
	26	8.00	4.47	7.40	3.28	3.60	2.19
	37	5.80	3.42	6.80	2.16	2.80	1.78
	48	6.60	3.36	7.40	1.94	2.80	1.78
z	14	6.60	3.64	4.80	2.50	6.00	2.73
	23	7.00	3.00	3.00	1.15	5.60	2.40
	58	5.60	2.50	2.80	1.09	5.60	2.88
	67	7.60	2.07	2.80	1.09	5.20	2.58
xy	16	8.60	2.86	8.00	1.73	6.40	3.36
	25	8.00	3.08	6.40	1.94	6.00	3.39
	38	7.00	2.54	7.80	1.92	6.20	3.19
	47	7.20	2.48	6.60	2.19	5.40	3.28
xz	13	7.00	3.24	6.20	3.00	6.00	2.34
	24	7.60	3.50	4.80	2.16	7.00	2.73
	57	8.00	0.00	3.60	1.34	7.00	2.00
	68	9.20	1.48	3.80	1.09	7.20	2.77
yz	18	6.40	3.36	5.40	1.94	8.30	2.28
	27	6.60	3.04	9.00	1.87	8.00	3.67
	36	8.80	2.48	8.40	1.94	8.60	2.50
	45	8.80	2.86	8.80	0.09	8.80	2.68
xyz	28	7.60	3.04	8.80	1.64	9.80	1.30
	35	9.80	1.09	9.00	1.87	9.80	1.30
	46	9.80	1.30	8.40	1.81	9.60	1.34
	17	7.20	2.88	5.40	2.30	9.80	1.30

All three stimulus sets were once again subjected to multidimensional scaling. The metric solutions for the three stimulus sets appear in Figs. 1, 2, and 3. An examination of these figures indicates that the solution generally conforms to previous scaling solutions with abstract stimulus sets, particularly for the harmonic oscillator problem and the combined stimulus set. In the case of the falling body prototype problem, as might be inferred from an examination of the data in Table 2, there is a good deal more distortion than we have found in any previous figure.

Mean response, latency, and standard deviation were also computed and these appear in Table 3. Examination of the data here reveals general comparability between the three stimulus sets. In general, subjects took much longer to respond to the prototype problems than they did to respond to the abstract stimulus set presented in Chapters 3 and 4. Mean response latency was also computed across all stimulus sets and is presented by trials in Fig. 4. The general asymtotic decreas in latency as a function of trials is once again apparent, indicating that subjects become more proficient in their responses with increasing experience.

We conclude that subjects are generally able to detect distinctive features in the more complex stimulus set although, as might be expected, with increasing stimulus complexity the variability among stimulus pairs also increases. The most striking finding of an examination of the judgme ; data, however, is the fact that the response patterns which occurred for the simpler stimulus sets once again appear. This indicates that the model proposed in chapter 2 does indeed have considerable generality and can be used to interpret not only simplified alpha-numeric expressions that are virtually without semantic meaning, but also much more complicated expressions which derive their meaning from the manner in which they account for physical phenomena in the real world.

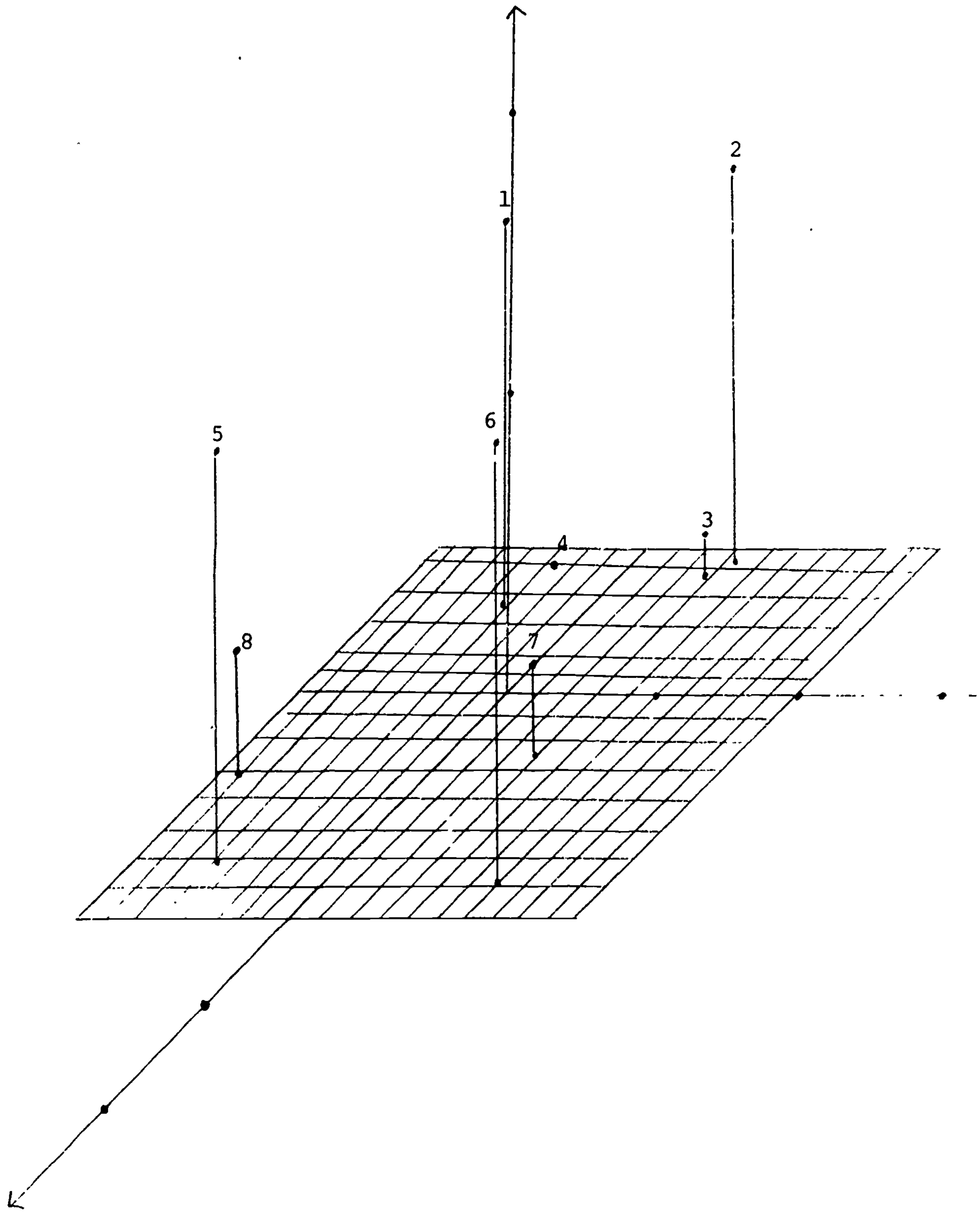


Fig. 1.Harmonic Oscillator, Metric Euclidean Scaling

Table 4

Mean Response Latencies and Standard Deviations for each Stimulus Pair on Last Trial

Category	Stimulus Pair	Harmonic Oscillator $\bar{X}$	Harmonic Oscillator s	Falling Body $\bar{X}$	Falling Body s	Combined $\bar{X}$	Combined s
x	12	4.83	2.91	5.57	1.56	5.12	1.59
	34	9.93	7.09	2.91	1.26	9.59	5.17
	56	11.06	6.60	6.53	2.15	5.91	1.31
	78	7.58	3.99	5.96	2.42	4.48	2.31
y	15	6.67	5.32	12.69	11.88	10.55	4.80
	26	5.88	3.53	12.22	7.82	9.93	4.69
	37	7.66	5.42	7.05	2.12	9.49	3.78
	48	7.58	3.99	5.54	1.27	6.31	2.46
z	14	5.76	3.10	4.80	2.63	8.82	1.64
	23	8.88	6.00	3.23	2.06	7.87	3.94
	58	8.28	5.00	5.85	0.69	4.84	2.17
	67	7.04	3.24	7.97	3.31	5.12	1.69
xy	16	6.51	5.07	13.51	6.88	10.77	3.65
	25	8.25	5.98	11.31	6.21	13.25	6.02
	38	10.41	10.44	7.99	3.88	7.93	1.48
	47	5.17	2.94	13.14	14.97	7.32	4.19
xz	13	7.85	5.83	6.37	3.54	10.72	6.36
	24	7.22	3.51	4.94	3.49	8.35	5.17
	57	16.80	14.94	9.41	3.46	5.56	1.43
	68	12.53	13.07	6.52	1.29	5.30	2.34
yz	18	6.81	4.05	11.26	5.47	9.79	4.25
	27	5.83	4.14	12.34	9.46	7.44	2.03
	36	9.94	8.09	14.74	15.67	10.39	2.05
	45	9.12	7.85	8.40	4.63	9.13	4.12
xyz	28	8.66	5.85	10.38	1.51	10.04	2.25
	35	7.43	7.34	15.15	5.18	8.01	2.42
	46	11.09	9.22	10.74	4.30	8.20	1.97
	17	6.77	5.71	7.15	2.39	9.69	1.77

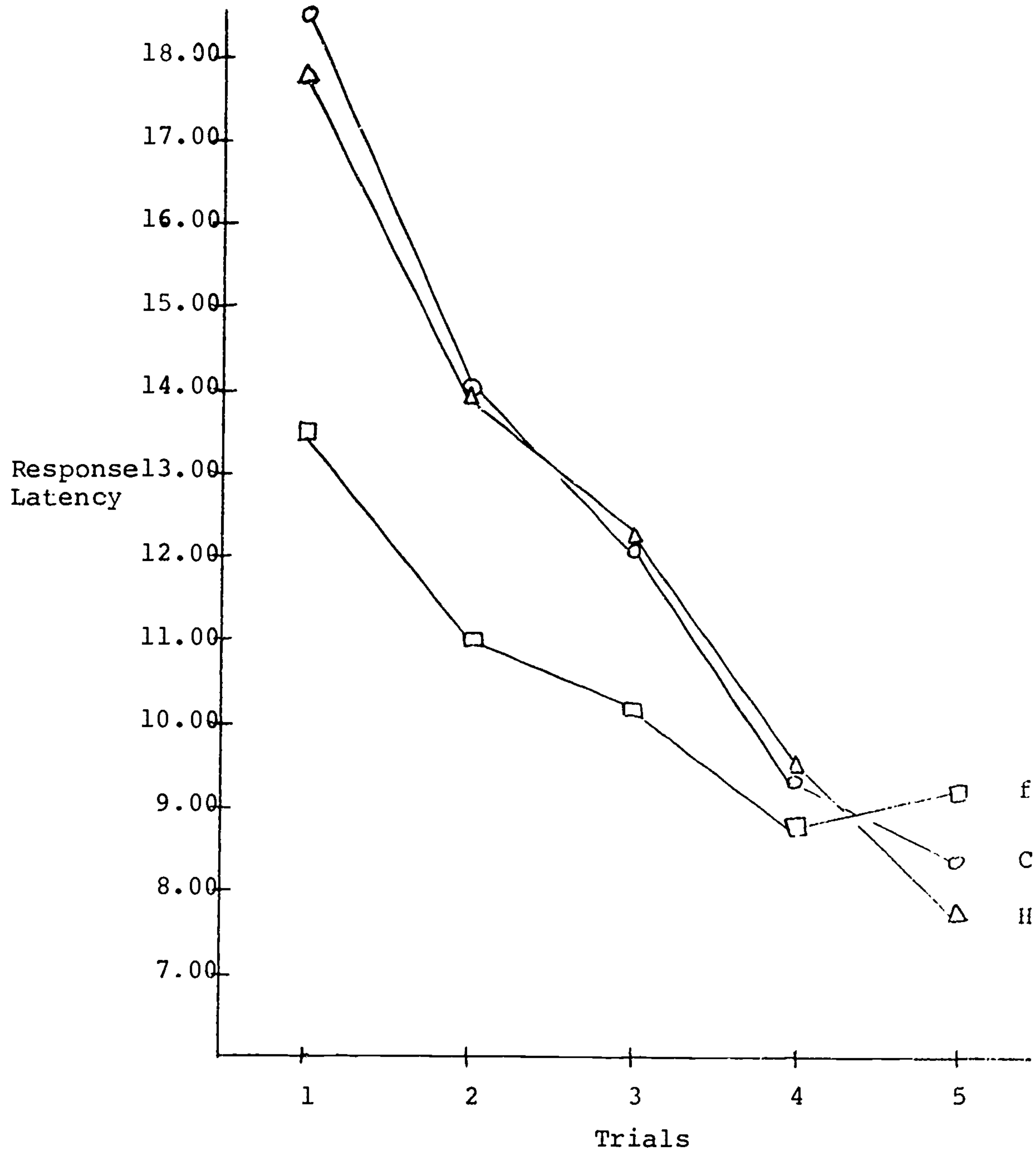


Fig. 4. Mean Response Latency by Trial (Averaged Over All Stimulus Pairs).

Encoding Task. Immediately after a S completed the dissimilarity judgments, he performed the encoding task. Materials consisted of the eight expressions depending on the task he had performed, printed one each on 2-1/2 x 3 inch cards.

E shuffled the stimulus cards and presented them in a haphazard array, approximately rectangular, for each trial. One expression was selected by E as a "target" and placed to Ss left of the remaining 7 cards. According to the target, S selected 3 of the remaining stimuli which made that target seem "odd" or inappropriate in the: · context. T-us, S constructs or encodes a message set of expressions, one being odd.

S was told that the optimal strategy would be to choose contrast sets such that the 3 expressions were all more similar to each other than any was to the target. S was allowed to manipulate the materials as he desired and identified his selections by handing the contrast set to E. No time limit was invoked.

Each of the eight expressions served in turn as a target according to a pre-arranged random order. Thus, each S produced eight message sets.

Session II -- Decoding Task

All fifteen of the Ss completed this portion of the experiment, all within one to fourteen days after the first session.

Each subject was given 32 message sets each consisting of four of the expressions from the task he had performed previously. The four expressions were typed on 5 x 5 inch cards and the subject was given a shuffled deck of 32 of these cards. S was reminded of the previous encoding task and was instructed to choose the expressions that seemed most odd in the set of four, and to go through the whole deck in this fashion.

Decoding sets were originally constructed on the basis of two paradigms. Due to an unfortunate error in the numbering of the stimuli which went undiscovered until the conclusion of the experiment, only one of these can be utilized in the analysis of decoding sets, though the other has been included in comparing encoding to decoding. This set, termed E-defined, since its relational characteristics derive from expressions in the abstract or formal model (cube), was constructed to reflect 3 nodes of one face connected via 2 diagonals and one superdiagonal

to one node of another face. This set was also expected to obtain from the Ss dissimilarity judgments.

Though message sets were essentially E-defined, they had the possibility of containing from 0-8 "own" sets, i.e., sets which corresponded to those which the S himself had encoded during the previous session.

Results

Encoding Task:

A relation between dissimilarity judgments and choices of contrast sets was confirmed. Those of optimal behavior (target -- contrast sets predicted by dissimilarity judgments) were made by examining the proportion of responses summed across all Ss and all targets falling into optimal (O), nonoptimal (N), and toss-up (T) categories.

Optimality was determined by rank sum test comparing the three target-contrast dissimilarities with the contrast-contrast dissimilarities. In order to be considered as O, the minimal stimulus rank sum must be 9 or less and at least 1.0 rank sum units from the nearest rank sum.

The row marginals of Table 5 demonstrate that the proportion of O, N, and T selections of contrast sets are .716, .141, and .141 respectively. Assuming the criteria established above, the probability of an expression having an O rank sum is .19 if there are no ties, thus predicting an expected number of 1.52 optimal sets of 8 or a .190 proportion. The observed range of projection for individual subjects was .200 to 1.000.

Assuming the same criteria, but summing across subjects for a given target, the observed proportion of O sets ranged from .666 in targets 1, 2, 5 and 8 to .800 on target 6.

The combined task obtained the highest proportion (.900), the falling body task next (.700) and the harmonic oscillator last (.555).

Session II -- Decoding

Decoding consistency: Consistency of decoding is evaluated by comparison of S's encoded sets with S's decoded sets. The proportion of encoded (own) sets consistently decoded was .700. The combined task obtained

Table 5

Proportion of Optimal, Non-optimal, and Toss-up Selections for Encoding Task

Task	S	Sum of Choices per S: O	N	T	Target Expressions: 1	2	3	4	5	6	7	8	p(O)	p(N)	p(T)
Harmonic Oscillator	1	6	1	1	O	N	O	O	O	N	O	T	.750	.125	.125
	3	5	3	-	N	O	N	O	O	O	O	N	.625	.375	--
	5	4	1	3	O	T	O	O	N	T	T	O	.500	.375	.125
	10	4	3	1	N	O	O	N	T	O	O	N	.500	.375	.125
	12	3	1	4	T	N	T	T	T	O	O	O	.375	.125	.500
Falling Body	2	5	1	2	T	T	O	N	O	O	O	O	.625	.125	.250
	4	6	-	2	T	O	O	O	O	T	O	O	.750	--	.250
	8	8	-	-	O	O	O	O	O	O	O	O	1.000	--	--
	13	4	3	1	O	O	O	O	T	N	N	N	.500	.375	.125
	14	5	2	1	O	N	N	O	O	O	T	O	.625	.250	.125
Combined Task	6	7	1	-	O	O	O	O	O	O	O	N	.875	.125	--
	7	6	1	1	O	O	O	O	N	O	T	O	.750	.125	.125
	9	8	-	-	O	O	O	O	O	O	O	O	1.000	--	--
	11	8	-	-	O	O	O	O	O	O	O	O	1.000	--	--
	15	7	-	1	O	O	T	O	O	O	O	O	.875	--	.125
	N=15	86	17	17									.716	.141	.141

Proportions of O, N, T on Each Target Expression

	1	2	3	4	5	6	7	8	
O	O=10 p(O)=.666	10 .666	11 .733	12 .800	10 .666	11 .733	11 .733	10 .666	86
N	N=2 p(N)=.133	3 .200	2 .133	2 .133	2 .133	2 .133	1 .066	4 .266	17
T	T=3 p(T)=.200	2 .133	2 .133	1 .066	3 .200	2 .133	3 .200	1 .066	17

Total Proportions of O, N, T on Each Task

	O	N	T	p(O)	p(N)	p(T)
Harmonic Oscillator	22	9	9	.555	.225	.225
Falling Body	28	6	6	.700	.150	.150
Combined Task	36	2	2	.900	.050	.050
	86	17	17	.716	.141	.141

Table 6

Proportion of Optimal, Non-optimal, and Toss-up Choices for "Own" Sets

Task	S	Sum of Choices per S: O	N	T	Target Expressions: 1	2	3	4	5	6	7	8	p(O)	p(N)	p(T)
Harmonic Oscillator	1	4	-	1	O	-	-	T	-	O	O	O	.800	--	.200
	3	1	-	-	-	-	-	-	-	-	O	-	1.000	--	--
	5	1	-	-	-	-	-	-	-	O	-	-	1.000	--	--
	10	1	2	-	N	-	O	T	-	-	N	-	.250	.500	.250
	12	-	3	-	-	-	-	-	-	N	N	N	--	1.000	--
Falling Body	2	5	-	1	O	O	O	-	O	-	T	O	.867	--	.133
	4	1	4	-	-	N	-	O	-	N	N	N	.200	.800	--
	8	6	-	-	-	O	O	O	O	O	O	-	1.000	--	--
	13	1	1	1	-	O	N	-	-	-	T	-	.333	.333	.333
	14	-	4	-	-	-	-	N	N	N	N	-	--	1.000	--
Combined Task	6	5	-	-	O	O	O	-	-	-	O	O	1.000	--	--
	7	5	-	-	O	O	O	-	-	-	O	O	1.000	--	--
	9	1	-	-	-	-	-	-	-	-	O	-	1.000	--	--
	11	2	-	-	-	-	O	-	-	O	-	-	1.000	--	--
	15	4	-	-	-	O	O	-	-	-	O	O	1.000	--	--
	N=15	37	14	3									.685	.259	.076

Proportions of O, N, T on Each Target Expression

	1	2	3	4	5	6	7	8
O	4 .800	6 .867	7 .876	2 .400	1 .500	4 .572	7 .538	5 .715
N	1 .200	1 .133	1 .124	1 .200	1 .500	3 .428	4 .307	2 .285
T	0 .000	0 .000	0 .000	2 .400	00 .000	0 .000	2 .153	0 .000

Total Proportions of O, N, T on Each Task

	O	N	T	p(O)	p(N)	p(T)
Harmonic Oscillator	7	5	1	.538	.384	.076
Falling Body	13	9	2	.541	.375	.083
Combined Task	17	-	-	1.000	--	--
	36	14	3			

the highest accuracy (1.000), followed by the falling body (.708) and the harmonic oscillator (.421).

Decoding accuracy: Accuracy of decoding was evaluated with respect to accuracy of decoding message sets encoded by oneself (own), and E-defined sets.

Own sets: The proportion of own sets correctly decoded was .685 (Table 3). Again, the chance proportion of optimal success was .19. The observed range of optimal response for individual subjects was .000 (12) to 1.000. The proportion of O sets for any given encoded target was .400 (T4) to .876 (T3). The rank of targets according to proportion of optimality does not reveal any ordering that can be compared in relation to encoding vs. decoding. In the decoding task, the target order from most to least O:

3 > 2 > 1 > 8 > 6 > 7 > 5 > 4,

whereas for encoding:

4 > 3, 6, 7 > 1, 2, 5, 8.

Again, the combined task obtained the highest proportion of O (1.000) followed by the falling body (.541) and harmonic oscillator (.528).

E-defined: E-defined sets also contain some sets previously designated as "own" since decoding sets were only compared to encoding sets if they corresponded by chance. A subject could have from 0 - 8 own sets in the decoding task.

The E-defined category was predicted to reflect S's dissimilarity judgments. In this set the model clearly predicts which expressions within each set should be "odd" on the basis of distances between vertices in the model. By dissimilarity judgments the O proportion of this set was 237/349 or .679, while by the model the O proportion was 244/349 or .699 (Table 7). (In this analysis the data for subject 12 was omitted due to the fact that he was clearly responding in a totally inconsistent fashion). The combined task proved to be highest in both cases, .752 by dissimilarity judgments (dj's) and .832 by the model, followed by the harmonic oscillator, .663 by dj's and .673 by the model. The falling body task ranked lowest, .617 by dj's and .583 by the model.

The ranking of the combined task in this case is interesting. Note that there is not perfect consistency between own set encoded and own set decoded over all tasks, average consistency being .685. The decoding set by dissimilarity judgments is in a sense a consistency measure between the dj's of the first task and the oddity judgments of the second, and obtains a .679 proportion

Table 7

Decoding Task
Comparison of Optimal Choice Proportions on
E-Defined Sets
vs.
Model Prediction

	S	E-defined Set (Dissimilarity Judgments) p(O)	E-defined Set (Model Predicted) p(O)
harmonic oscillator	1	22/26	19/26
	3	15/26	15/26
	5	19/26	21/26
	10	13/26	15/26
	12	--	--
Falling body	2	19/24	14/24
	4	8/24	9/24
	8	22/24	18/24
	13	10/24	15/24
	14	15/24	14/24
Combined task	6	13/25	22/25
	7	17/25	17/25
	9	18/25	25/25
	11	23/25	20/25
	15	23/25	20/25
		237/349	244/349
		(p).679	(p).699

Total Proportions on Each Task

Task	E-set (DJs) p(O)		E-set (model predicted p(O)	
Harmonic Oscillator	69/104	(p).663	70/104	(p).673
Falling Body	74/120	(p).617	70/120	(p).583
Combined Task	94/125	(p).752	104/125	(p).832
Total	237/349	(p).679	244/349	(p).699

over all tasks. The combined task obtains 1.000 in encoding vs decoding consistency and .752 by dj's. This task is essentially an easier task than the others since it does not include the three unlabeled terms or the complexity of expressions. Since this task is the most stable of the three, it is reasonable to conclude that the difficulty of the task influences the stability of the structure indexed by the dissimilarity judgments.

In this context, it may also be asked why any instability occurs at all, i.e., why do subjects not respond perfectly in the decoding task according to the structure indexed by their judgments. No regularities were observed in the data that could answer this question. Inspection of decoding sets in each task which had the highest proportion of non-optimal choices by dj's (3 of 5 subjects making N choices) showed a high percentage of situations in which subjects made optimal choices by the model though the choice was non-optimal by dj's; 18/25 in the harmonic oscillator, 8/12 in the combined, and 11/18 in the falling body. Again, note that less of these situations occur in the combined task. This indicates that the communications task is indeed a different task from dissimilarity judgment. Though drawing on the same knowledge, the results of this operation appear to be different in each case, yet the correspondence between the judgments and communications task indicates their convergence on the same knowledge.

The communications task as a converging operation tends to reflect judgments by Ss consistent with (a) the subjects' dissimilarity judgments and (b) the logical model. Further, Ss tend to be consistent across the three tasks: dissimilarity judgments, encoding and decoding. Stability of the structures indexed by the dissimilarity judgments is indicated to be a function of the difficulty of the task itself.

Chapter 7

A Replication With Longitudinal Data

The purpose of this experiment was to gather longitudinal data on both the abstract and prototype problem stimulus sets. The experiment was designed to replicate the graduate physics data presented in chapter 3 and chapter 6 and the undergraduate data collected in chapter 5, as well as gather new data on undergraduates with respect to the prototype problems. If successful, the replication would demonstrate the stability of knowledge structures over time, and also show the convergence of student competence with respect to abstract and prototype stimuli.

The experiment was performed by undergraduate and graduate physics students. One group of undergraduates and one group of graduate students were exposed to the αβγ, mxt, and gm-cm abstract stimulus sets (the color set was omitted), a second group of undergraduates and a second group of graduate students received the combined stimulus set. Ss received 3 trials on the abstract stimuli and 5 trials on the combined stimuli. As before, the stimulus pairs were presented on each trial in a random order. However, the abstract stimulus sets were presented in a fixed order: αβγ, gm-cm and then mxt. This was done in order to obtain an accurate assessment of RL for the mxt set. In the previous procedure (described in chapter 3) some Ss had the mxt set first and some last so that RL's were averaged across different stages in task performance. Assuming Ss become proffICIENT between as well as within tasks, we decided to obtain as accurate an index of RL as possible in order to determine its future usefulness in this type of work.

Fall Testing

Ss were tested using the procedures outlined in chapters 3 and 6. Nine undergraduate and five graduate Ss were tested on the abstract stimuli and 10 undergraduate and 5 graduate Ss were tested on the prototype problem stimuli (combined set).

Results and Discussion

Following the standard format developed in chapter 3, intertrial correlations were computed for both response and response latency in the abstract stimulus set for each group of Ss. These data are presented in Table 1 and

Table 1

Intertrial Correlations for Response and Response Latency (RL)
Graduates and Undergraduates
Fall Testing

Response

Stimulus Set	Graduates Trials 1-2	Graduates Trials 1-3	Graduates Trials 2-3	Undergraduates Trials 1-2	Undergraduates Trials 1-3	Undergraduates Trials 2-3
αβγ	.82	.86	.92	.85	.93	.95
mxt	.93	.95	.95	.88	.96	.97
gm-cm	.85	.88	.94	.91	.96	.95
Response Latency						
αβγ	-.08	-.09	.69	.46	.11	.20
mxt	.20	.34	.78	.45	.58	.18
gm-cm	.12	.11	.32	.39	.35	.49

reveal much the same thing we have found before. Namely, that correlations between adjacent trials are higher than for nonadjacent trials and all correlations are higher for response data than for latencies.

Once again, subsequent analysis was performed on trial 3 data. Mean responses and standard deviations for each stimulus are presented in Table 2 for the graduate students and Table 3 for the undergraduates. Mean latencies together with their standard deviations appear in Table 4 for graduates and Table 5 for undergraduates. Comparing graduate and undergraduate data for both response and response latency, the most apparent difference is that stimulus pairs are judged more dissimilar by graduates than undergraduates and the undergraduates tend to take more time in making their judgments. An inspection of the data indicates that the effect of task order is not large, although as can be seen by comparing the latencies in Tables 4 and 5 with the latencies in Table 6 of chapter 3, placing a set last results in lower latencies than when the sets occur in a random order.

Once again, the mean judgments in each set were subjected to multidimensional scaling procedures. And as before, the configurations were regular polyhedrons in 3 dimension with low stress much like the configurations in chapter 3. The Fall testing with abstract stimuli can be viewed simply as a replication of our previous findings.

Responses in the combined task were next analyzed to determine how the two S groups differed when the stimulus pairs involved semantic content. The intertrial correlations on the combined stimulus set for both graduate and undergraduate students appear in Table 6. These correlations are consistent with data collected on the combined stimulus set with graduate students as described in chapter 6.

Mean dissimilarity ratings on the combined stimulus set appear in Table 7 for the graduate students and undergraduate Ss. Both the variability and the magnitude of students' judgments are greater for graduates than for undergraduates (as was the case with stimuli from the abstract set). Response latency data for the two groups of subjects appears in Table 8. Graduate students seem to take more time to make their judgments than undergraduates, though the difference between the two groups is less pronounced on the combined stimulus set than it was for the abstract set.

Table 2

Mean Dissimilarity Rating and Standard Deviation for Each Stimulus Pair on Last Trial (Fall Testing)
Graduate Students

Category	Stimulus Pair	αβγ		mxt		gm-cm	
		x $\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
x	12	2.20	.45	3.40	2.61	5.20	3.77
	34	3.00	1.22	3.20	2.17	5.20	3.77
	56	3.75	2.87	3.40	2.61	6.40	3.44
	78	2.00	.82	3.20	2.17	5.20	3.77
y	51	6.40	3.78	7.00	3.08	6.40	3.78
	62	7.00	3.74	7.20	3.19	6.80	2.95
	73	9100	1.41	7.00	3.81	6.60	2.97
	84	8.00	3.46	8.00	2.55	5.80	3.03
z	14	7.00	2.83	6.40	3.21	5.40	3.13
	23	7.25	2.22	6.60	2.97	6.00	2.92
	58	8.00	2.24	7.40	2.61	6.40	3.78
	67	8.50	1.29	7.80	2.86	7.00	3.32
xy	16	8.80	1.92	8.60	2.30	8.50	2.65
	25	8.40	2.19	9.40	1.14	9.80	1.30
	38	7.80	2.59	7.80	3.27	8.20	2.77
	47	6.60	3.91	7.80	2.39	8.40	2.51
xz	31	7.75	2.22	8.00	2.35	8.40	2.51
	42	7.75	2.06	8.20	2.17	8.40	2.88
	75	7.80	2.68	7.80	2.77	9.00	2.55
	86	8.50	2.38	9.00	.82	8.75	2.22
yz	18	7.40	3.65	7.40	4.10	5.20	3.27
	27	7.20	4.44	6.40	4.16	7.00	2.92
	36	10.00	1.00	10.00	1.00	8.20	2.39
	45	10.20	.84	10.00	.71	7.80	2.86
xyz	53	10.50	.58	10.20	.84	9.80	1.30
	64	10.50	.58	10.20	.84	9.80	1.30
	71	7.25	4.50	7.40	4.28	8.40	2.51
	82	6.60	4.04	8.60	2.61	9.00	2.12

Table 3

Mean Dissimilarity Rating and Standard Deviation
for Each Stimulus Pair on Last Trial (Fall Testing)
Undergraduates

Category	Stimulus Pair	αβγ		mxt		gm-cm	
		$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
x	12	2.78	2.05	2.44	.53	3.11	1.62
	34	2.62	.74	2.78	1.09	3.75	1.75
	56	2.33	.50	2.78	1.09	3.22	1.56
	78	2.25	1.16	3.22	1.92	3.44	1.88
y	51	4.89	2.67	5.22	2.68	4.89	3.02
	62	4.00	2.18	5.56	3.13	5.44	3.24
	73	6.22	3.11	4.78	2.77	4.44	2.24
	84	4.56	2.92	5.78	2.82	4.67	2.74
z	14	5.11	3.10	4.44	1.88	3.88	2.85
	23	5.67	2.40	5.11	2.47	4.44	2.79
	58	4.89	2.32	6.50	2.45	5.11	3.10
	67	6.89	2.76	6.22	2.64	5.78	3.23
xy	16	6.11	2.98	6.56	2.70	7.00	3.16
	25	5.44	2.18	6.78	2.82	6.11	2.71
	38	6.62	2.62	6.25	2.71	5.67	2.40
	47	5.89	3.33	5.67	2.40	6.00	2.55
xz	31	6.11	2.52	6.56	2.46	5.33	2.12
	42	7.11	2.80	6.44	2.[illegible]6	6.11	2.32
	75	6.11	2.85	7.00	2.50	7.11	3.02
	86	6.89	2.47	6.44	2.55	6.25	2.66
yz	18	5.89	2.89	5.44	3.32	4.78	2.77
	27	7.22	3.83	5.33	2.96	5.56	2.46
	36	6.33	2.64	7.22	2.11	7.67	2.96
	45	7.56	2.55	7.78	2.44	6.67	2.45
xyz	53	7.78	2.11	6.89	2.90	8.11	2.26
	64	8.22	2.54	7.00	2.24	7.78	2.22
	71	6.33	3.39	7.22	3.14	7.44	2.65
	82	5.56	3.24	8.00	3.20	7.00	2.45

Table 4

Mean Response Latency and Standard Deviation
for Each Stimulus Pair on Last Trial (Fall Testing)
Graduate Students

Category	Stimulus Pair	αβγ $\bar{X}$	αβγ s	mxt	mxt s	gm-cm $\bar{X}$	gm-cm s
x	12	2.37	.66	2.11	.94	2.15	.70
	34	3.16	.67	2.23	.80	2.93	1.31
	56	2.84	.70	3.44	1.28	3.09	.64
	78	3.48	1.46	3.20	1.56	2.57	.76
y	51	2.97	1.54	3.42	1.96	2.14	.45
	62	2.47	.62	2.79	.66	3.11	.26
	73	4.03	1.77	3.43	1.41	3.91	1.57
	84	2.86	1.45	3.08	1.69	2.76	1.15
z	14	2.33	.78	2.24	.55	2.46	.91
	23	3.55	1.26	2.15	.28	3.32	.98
	58	3.26	1.00	3.38	.59	2.52	.61
	67	4.48	1.82	4.84	2.74	3.77	.60
xy	16	5.95	6.52	3.33	1.17	3.53	1.67
	25	3.56	1.45	3.06	.92	2.59	.95
	38	4.66	1.73	3.87	2.79	2.78	1.28
	47	3.53	1.34	3.44	1.39	2.70	1.19
xz	31	3.18	1.59	3.51	1.91	2.81	1.10
	42	3.39	1.33	2.75	1.27	2.83	.84
	75	5.68	3.08	3.70	.61	2.47	.72
	86	2.78	.50	4.00	1.58	3.11	.91
yz	18	4.32	2.64	3.04	1.32	2.58	.87
	27	4.25	1.39	5.86	3.66	3.04	1.16
	36	4.02	2.02	3.65	1.37	2.84	.72
	45	3.21	1.89	2.71	1.14	2.26	.45
xyz	53	2.59	74	2.86	.86	2.21	.81
	64	2.83	1.22	4.17	1.98	2.82	1.03
	71	3.51	2.58	3.87	2.50	2.64	1.35
	82	3.22	1.23	4.38	4.00	3.91	1.71

Table 5

Mean Response Latency and Standard Deviation for Each Stimulus Pair on Last Trial (Fall Testing) Undergraduates

Category	Stimulus Pair	αβγ $\bar{X}$	αβγ s	mxt $\bar{X}$	mxt s	gm-cm $\bar{X}$	gm-cm s
x	12	3.73	1.46	3.60	2.02	3.36	1.31
	34	9.06	4.24	4.17	1.46	5.31	2.84
	56	7.12	4.27	12.03	14.24	5.75	3.40
	78	8.29	3.78	5.70	2.27	5.20	2.75
y	51	4.82	1.87	5.00	3.49	7.97	8.32
	62	4.76	1.34	14.67	22.17	7.24	5.96
	73	7.15	4.07	5.75	2.36	5.30	1.81
	84	16.39	26.90	6.35	2.29	4.46	1.52
z	14	3.10	1.10	4.20	2.29	2.85	.77
	23	6.13	2.25	4.47	1.22	4.52	1.36
	58	6.54	3.30	6.07	1.72	4.62	2.27
	67	8.63	2.98	6.02	2.01	6.20	4.24
xy	16	5.76	2.01	12.70	18.87	4.52	.90
	25	7.17	3.60	9.21	8.29	5.88	3.87
	38	17.25	22.32	8.00	4.11	8.58	5.53
	47	7.00	5.44	6.62	2.09	8.35	8.75
xz	31	6.08	3.23	5.82	3.70	6.40	2.86
	42	9.74	13.06	4.62	1.63	3.92	.69
	75	6.02	2.02	7.45	4.36	9.15	7.83
	86	8.07	2.21	6.63	1.91	6.32	1.40
yz	18	6.61	2.18	8.30	4.42	3.66	.94
	27	7.41	3.19	7.09	3.03	7.36	3.90
	36	7.34	2.83	6.16	1.91	5.93	2.93
	45	8.94	4.84	6.34	2.11	5.74	3.72
xyz	53	6.88	2.32	6.37	2.69	7.66	2.46
	64	4.82	1.47	6.37	3.28	6.32	3.16
	71	10.67	6.81	6.64	2.97	5.19	3.89
	82	9.32	4.92	7.73	4.52	6.98	5.25

Table 6

Intertrial Correlations for Response and Response Latency (RL) on Combined Set (Graduates and Undergraduates) Fall Testing

	Response		Response Latency	
Trial	Graduates	Undergraduates	Graduates	Undergraduates
1-2	.79	.95	.30	.13
1-3	.78	.95	.40	.43
1-4	.69	.91	.23	.12
1-5	.77	.94	.32	.38
2-3	.89	.96	.58	.26
2-4	.87	.96	.39	.48
2-5	.89	.95	.38	.10
3-4	.89	.94	.34	.16
3-5	.91	.97	.67	.24
4-5	.95	.94	.43	.31

Table 7

Mean Dissimilarity Rating and Standard Deviation for Each Stimulus Pair in the Combined Set on Last Trial (Fall Testing)

Category	Stimulus Pair	Graduates $\bar{X}$	Graduates s	Undergraduates $\bar{X}$	Undergraduates s
x	12	7.20	4.09	4.40	2.12
	34	4.80	3.56	3.80	2.20
	56	3.40	4.34	2.67	3.04
	78	4.00	1.87	3.80	2.74
y	51	7.00	3.32	6.10	2.73
	62	7.20	3.70	6.20	3.22
	73	4.40	3.78	6.40	3.69
	84	4.20	2.77	6.50	3.54
z	14	7.20	3.56	5.20	2.30
	23	5.80	3.56	4.60	2.84
	58	4.40	1.95	4.20	1.87
	67	3.60	1.14	3.60	2.22
xy	16	9.80	1.30	6.60	2.67
	25	7.60	3.05	6.90	3.07
	38	7.00	3.39	7.90	2.47
	47	7.40	3.78	7.80	2.97
xz	31	8.20	3.11	6.00	2.11
	42	6.60	3.36	5.40	2.22
	75	4.60	2.19	4.50	2.80
	86	5.40	3.36	4.90	2.51
yz	18	10.00	1.22	6.50	2.95
	27	8.00	2.55	6.80	2.15
	36	7.60	4.22	7.40	3.24
	45	7.00	3.54	8.30	2.95
xyz	53	7.60	3.78	7.80	3.49
	64	8.20	3.90	8.90	2.13
	71	8.40	2.61	7.90	2.18
	82	9.40	2.30	7.40	2.12

Table 8

Mean Response Latency and Standard Deviation for Each Stimulus Pair in the Combined Set on Last Trial (Fall Testing)

Category	Stimulus Pair	Graduates $\bar{X}$	Graduates s	Undergraduates $\bar{X}$	Undergraduates s
x	12	4.31	1.42	5.09	2.72
	34	2.75	.87	4.21	3.38
	56	4.99	3.92	5.49	3.96
	78	3.73	1.25	5.75	4.20
y	51	7.69	3.83	4.38	2.11
	62	5.48	2.66	4.59	2.86
	73	5.93	4.19	3.72	2.12
	84	5.23	2.80	4.86	4.82
z	14	3.26	1.57	3.52	1.88
	23	4.44	4.12	2.85	1.13
	58	5.49	2.32	6.94	5.85
	67	4.17	1.09	4.61	2.29
xy	16	7.58	3.35	5.29	5.23
	25	9.22	4.64	4.48	2.31
	38	6.66	3.37	3.53	1.89
	47	4.83	2.01	4.16	2.44
xz	31	4.97	2.58	6.19	5.36
	42	8.44	11.30	5.19	4.08
	75	4.34	2.16	4.53	2.52
	86	4.32	2.25	5.16	2.30
yz	18	6.69	2.73	6.43	8.79
	27	7.22	3.08	7.23	9.23
	36	6.35	1.91	3.66	1.74
	45	7.03	3.34	4.13	3.27
xyz	53	5.68	2.40	4.65	2.95
	64	7.64	3.21	5.13	4.25
	71	6.82	.62	4.26	2.86
	82	7.48	2.77	5.34	3.27

In general, the results of Fall testing with graduate and undergraduate students on the combined stimulus set reveals that each group of subjects responded to distinctive features of the stimulus set consistent with the placement of these features in our model for a structure of knowledge in physics.

Spring Testing

Eight undergraduates and five graduate students were tested with the abstract stimulus set. Five undergraduates and five graduate students were tested on the combined set. In each case, students had performed on the appropriate stimulus set in the Fall testing.

Intertrial correlations for both response and response latency were computed and are presented for the abstract stimulus set in Table 9. Mean judgments for the abstract stimuli are presented in Table 10 for the graduate students and Table 11 for the undergraduates. It appears that the judgments for the two groups of Ss are more alike than they were in the Fall testing with the possible exception of the gm-cm set. Response latencies were computed on the abstract stimulus sets and these appear in Tables 12 and 13. As in the Fall, undergraduates took more time to make their judgment than graduate students. (Compare these findings with the data in Fig. 7, chapter 5, which shows just the reverse).

Multidimensional scaling was performed on the mxt stimulus set for both graduate and undergraduate students to display these results under conditions of repeated testing. Fig. 1 contains the scaling solution for the graduate students and Fig. 2 contains the scaling solution for the undergraduate students. As might be expected from an examination of the mean dissimilarity judgments, the scaling solutions for the two groups of subjects are much alike (with some small, but noticable, differences).

The most significant finding from our analysis of the data from the abstract set is probably the fact that the effect of task order upon response latency is related to the degree of subject matter mastery. Thus, when the stimulus sets were administered in a random order (as in chapters 3 and 5) it appeared that graduate students took more time to make their judgments than undergraduates. However, when the task order is fixed (e.g., in this experiment $\alpha\beta\gamma$, gm-cm, mxt) then undergraduates take more time than graduate students on the latter tasks. The fact that randomization procedures can mask differences among Ss in processing time should be kept in mind for any future work attempted in this area.

Table 9

Intertrial Correlations for Response and Response Latency (RL)
Graduates and Undergraduates
(Spring Testing)

Response

Stimulus Set	Graduates Trials 1-2	Graduates Trials 1-3	Graduates Trials 2-3	Undergraduates Trials 1-2	Undergraduates Trials 1-3	Undergraduates Trials 2-3
αβγ	.87	.89	.95	.94	.95	.96
mxt	.98	.94	.93	.98	.98	.97
gm-cm	.97	.95	.97	.96	.96	.95
Response Latency						
αβγ	.18	-.01	.27	.45	.68	.72
mxt	.81	.55	.43	.21	.06	.58
gm-cm	.12	.22	.27	.37	.33	.32

Table 10

Mean Dissimilarity Rating and Standard Deviation for Each Stimulus Pair on Last Trial (Spring Testing) Graduate Students

Category	Stimulus Pair	αβγ $\bar{X}$	αβγ s	mxt $\bar{X}$	mxt s	gm-cm $\bar{X}$	gm-cm s
x	12	2.80	1.79	1.20	.45	4.00	3.94
	34	2.80	1.79	2.00	.00	5.40	3.71
	56	3.00	1.73	2.20	.45	6.00	3.67
	78	2.60	2.30	2.00	.71	5.00	3.67
y	51	4.80	1.30	7.40	2.07	5.60	4.04
	62	4.80	1.30	6.00	3.08	5.60	4.04
	73	5.00	1.00	4.75	1.26	6.00	3.39
	84	4.80	1.30	5.80	2.59	6.00	3.00
z	14	5.50	2.52	5.60	2.61	4.60	3.21
	23	5.20	2.49	5.40	2.30	4.80	3.27
	58	5.75	2.50	5.60	1.67	5.60	4.04
	67	5.40	.89	5.20	1.48	5.60	4.04
xy	16	6.80	2.17	5.20	1.92	8.20	2.77
	25	6.80	2.17	6.40	2.30	8.80	2.77
	38	6.80	2.17	6.25	1.71	8.20	2.68
	47	6.40	2.19	7.00	2.24	7.80	2.77
xz	31	7.00	2.83	5.60	1.14	7.80	2.77
	42	6.50	2.30	5.40	1.14	8.00	2.92
	75	7.20	2.95	6.80	1.92	8.40	2.88
	86	7.40	2.30	7.20	2.17	8.40	2.88
yz	18	6.00	3.81	5.80	3.83	5.20	2.86
	27	5.80	3.63	6.40	4.16	5.20	2.77
	36	9.80	.84	8.50	2.08	7.00	3.08
	45	8.40	2.07	9.00	2.00	6.80	3.11
xyz	53	9.20	1.79	9.75	1.26	8.40	2.34
	64	9.80	.45	9.80	1.10	10.00	1.58
	71	6.00	3.94	6.40	3.65	8.00	2.70
	82	6.40	4.10	6.60	4.22	9.00	.82

Table 11

Mean Dissimilarity Rating and Standard Deviation
for Each Stimulus Pair on Last Trial (Spring Testing)
Undergraduates

Category	Stimulus Pair	αβγ $\bar{X}$	αβγ s	mxt $\bar{X}$	mxt s	gm-cm $\bar{X}$	gm-cm s
x	12	2.38	.52	1.25	.71	3.62	2.13
	34	3.12	2.53	2.62	.74	3.38	1.30
	56	2.62	1.19	2.75	1.04	3.57	1.62
	78	1.75	.89	2.50	.53	3.88	1.88
y	51	5.00	1.41	5.25	2.05	3.75	1.67
	62	4.88	1.46	6.62	2.26	4.62	2.39
	73	5.62	2.56	5.86	1.68	4.12	1.46
	84	4.62	1.60	4.75	1.75	4.12	2.17
z	14	4.88	1.46	5.12	1.46	4.25	2.38
	23	5.00	1.41	5.12	1.46	4.88	2.36
	58	6.25	1.67	5.25	1.39	4.75	2.60
	67	6.00	1.77	5.88	2.59	5.38	1.68
xy	16	6.86	1.68	6.25	1.39	6.50	2.39
	25	6.38	1.60	6.38	1.30	5.62	2.13
	38	6.50	2.20	5.12	1.36	6.50	1.85
	47	6.00	2.39	6.62	2.62	6.12	1.55
xz	31	6.38	1.85	6.25	1.98	6.25	1.83
	42	6.88	1.46	6.25	1.98	6.50	1.60
	75	6.62	2.50	7.62	1.50	6.62	2.13
	86	6.88	2.36	7.28	1.70	6.38	2.20
yz	18	5.50	3.62	5.25	3.28	4.62	2.88
	27	5.38	3.42	5.62	4.00	6.50	2.20
	36	8.12	2.42	7.88	1.64	6.75	2.12
	45	8.12	1.88	8.75	2.25	6.62	3.02
xyz	53	9.62	2.06	9.43	2.15	7.75	2.25
	64	9.62	2.00	9.38	2.00	8.38	2.26
	71	5.88	4.29	5.62	3.50	7.38	2.20
	82	6.62	3.66	5.12	3.56	7.50	1.93

Table 12

Mean Response Latency and Standard Deviation for Each Stimulus Pair on Last Trial (Spring Testing) Graduate Students

Category	Stimulus Pair	αβγ $\overline{X}$	αβγ s	mxt $\overline{X}$	mxt s	gm-cm $\overline{X}$	gm-cm s
x	12	2.94	1.82	1.62	.25	2.34	1.72
	34	2.91	.78	2.42	.66	2.96	1.52
	56	3.13	1.08	3.22	.85	3.25	1.90
	78	3.87	1.49	3.16	.48	2.40	1.15
y	51	2.57	1.08	3.33	2.13	2.60	1.06
	62	3.86	2.14	2.78	1.05	2.56	.99
	73	3.40	1.26	3.32	1.48	2.98	1.65
	84	2.83	.82	2.52	1.16	1.89	1.01
z	14	2.29	.40	2.51	.88	2.57	1.26
	23	3.20	1.01	3.31	.41	3.12	1.02
	58	2.39	.73	3.11	1.52	2.08	.52
	67	5.18	.68	2.77	.60	2.94	.95
xy	16	3.63	1.55	2.48	.76	2.90	1.59
	25	3.93	1.11	2.37	.78	2.37	1.02
	38	3.78	.20	4.08	1.49	3.40	2.76
	47	3.01	.44	3.75	1.66	3.62	2.46
xz	31	2.96	1.41	2.80	1.38	2.90	2.50
	42	3.00	1.31	2.47	.75	2.86	1.50
	75	2.99	1.21	2.98	1.63	3.05	1.86
	86	4.04	1.39	3.74	1.89	2.52	1.31
yz	18	3.40	.91	3.58	1.81	2.81	1.29
	27	3.86	1.30	3.31	1.68	4.05	2.70
	36	3.80	1.58	3.53	1.08	3.28	.98
	45	3.15	.68	3.22	1.42	3.20	1.41
xyz	53	3.36	.95	3.62	.61	2.51	1.58
	64	3.62	1.26	3.76	2.01	2.24	1.51
	71	3.04	1.33	5.40	3.00	3.84	2.58
	82	5.07	2.53	2.84	.82	3.25	2.50

Table 13

Mean Response Latency and Standard Deviation for Each Stimulus Pair on Last Trial (Spring Testing) Undergraduates

Category	Stimulus Pair	αβγ $\overline{X}$	αβγ s	mxt $\overline{X}$	mxt s	gm-cm $\overline{X}$	gm-cm s
x	12	3.28	2.23	2.12	.93	3.55	1.45
	34	3.79	1.33	3.28	1.39	4.58	2.23
	56	5.67	1.37	5.36	3.61	6.02	5.59
	78	4.90	2.46	5.66	3.54	4.97	4.55
y	51	3.75	.85	4.39	2.92	3.10	.97
	62	5.46	1.16	4.70	1.80	4.38	2.54
	73	6.54	2.79	5.70	1.42	5.62	3.20
	84	4.46	1.34	3.60	.93	5.22	2.93
z	14	4.70	3.10	3.66	1.87	3.51	1.49
	23	4.13	1.32	4.36	1.69	4.39	2.91
	58	5.25	2.40	5.11	2.32	3.70	2.29
	67	5.82	2.47	5.33	2.03	7.41	8.39
xy	16	4.61	1.19	5.52	2.85	6.01	5.18
	25	5.58	1.15	3.87	.99	7.27	6.64
	38	5.14	2.09	7.53	5.17	4.07	.95
	47	5.70	2.16	4.35	1.85	6.14	5.51
xz	31	4.77	1.97	4.58	1.20	6.93	5.37
	42	4.14	1.88	6.59	3.60	8.02	10.41
	75	6.26	2.88	8.23	5.61	4.56	2.44
	86	6.80	4.53	9.75	6.30	5.26	2.34
yz	18	5.13	1.26	5.58	2.00	3.62	1.72
	27	6.85	2.64	5.78	2.34	5.34	2.44
	36	5.33	2.43	5.18	1.86	5.47	3.28
	45	5.32	1.70	5.66	3.08	4.48	2.56
xyz	53	5.53	3.11	5.55	1.22	6.84	3.82
	64	7.49	5.28	5.28	3.70	4.82	1.57
	71	5.24	1.80	5.95	2.66	5.99	4.38
	82	4.77	1.52	5.77	3.20	5.58	3.31

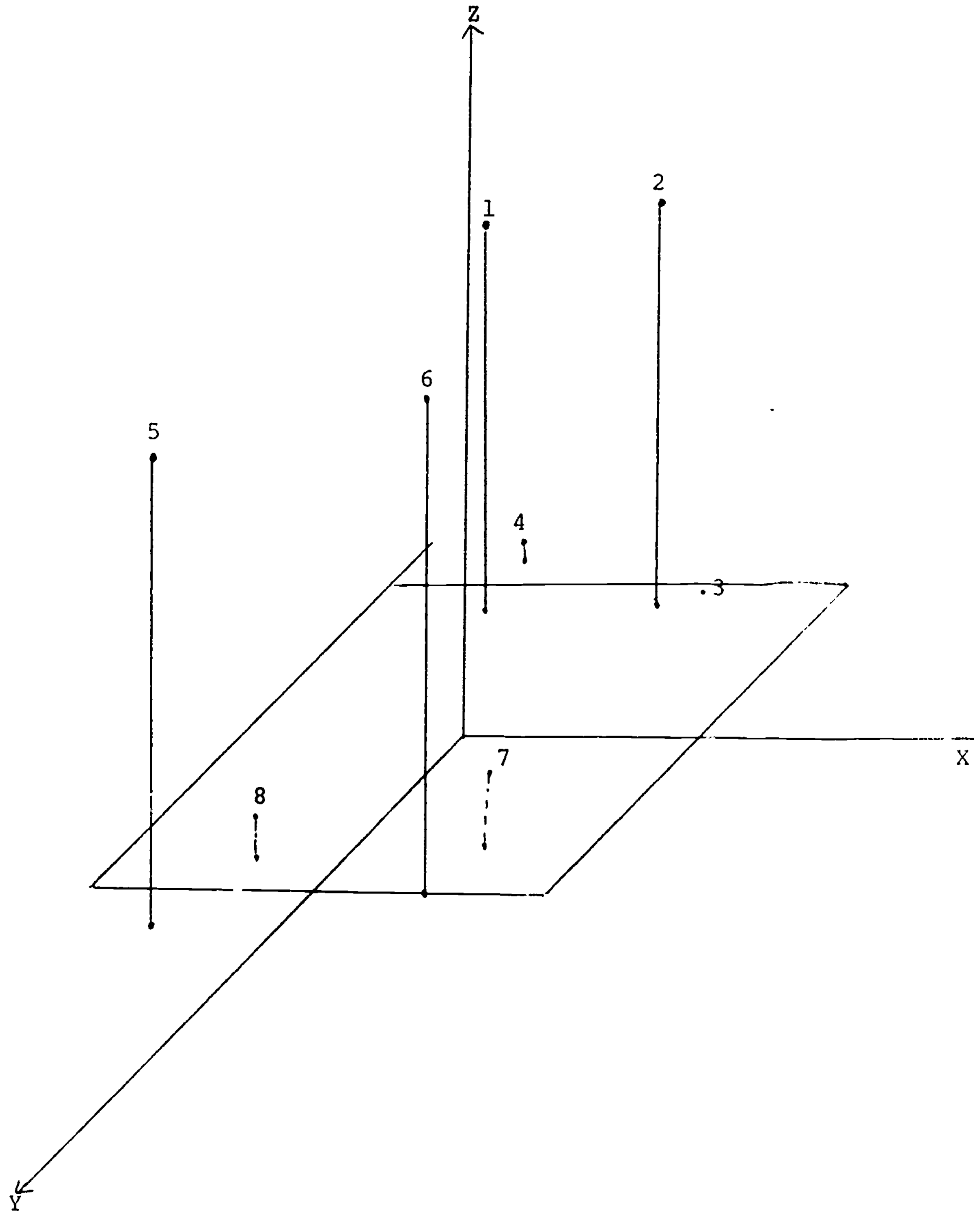


Fig. 1. Graduate Students (Spring Testing), MXT, Metric Euclidean.

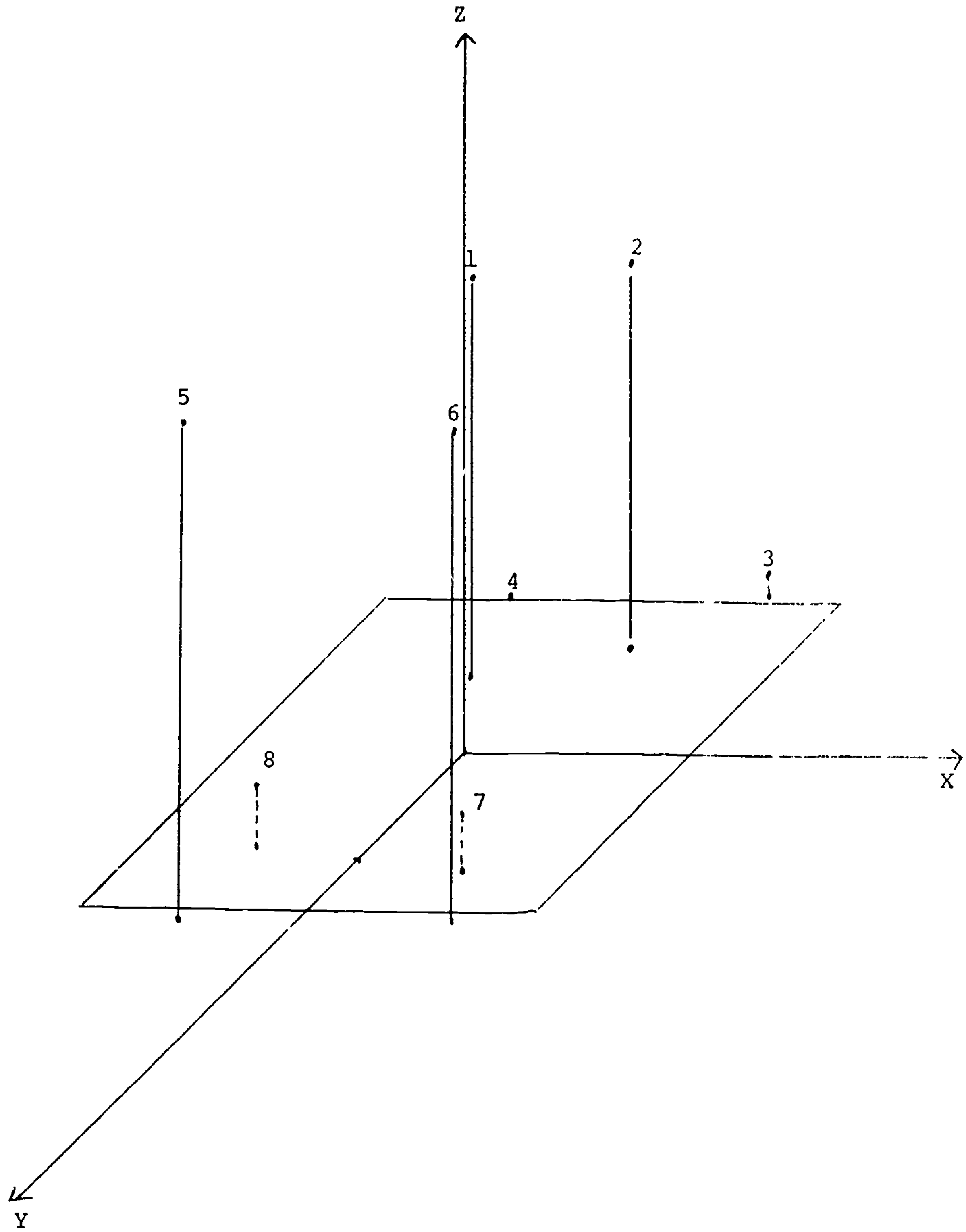


Fig. 2. Undergraduates (Spring Testing), MXT, Metric Euclidean.

Data from the combined set were next analyzed intertrial correlations for both response and response latency appear in Table 14. Mean dissimilarity judgments on the combined set are presented in Table 15 and mean latencies are presented in Table 16. Both graduate and undergraduate students once again reveal that stimulus similarity is judged to be an increasing monotonic function of the number of transformations that can be used to account for the differences between members of a stimulus pair. As in the Fall testing, graduate students require slightly more time to make their judgments than do the undergraduates. Whether the difference in latency is the opposite of what it was for the abstract set cannot be determined since Ss on the combined set had no prior practice sets (such as the αβγ set). The findings for the combined set as regards R_L are consistent, however, with the data presented in chapter 5.

Multidimensional scaling solutions were computed for the combined stimulus set and these are presented in Figs. 3 and 4. An examination of these figures reveals a greater difference between solutions for graduates and undergraduates than was the case with the abstract stimulus set. In general each solution can be viewed as a distortion of the model presented in Fig. 3 of chpater 2. However, the nature of the distortion differs depending upon whether one is an undergraduate or graduate student in the subject matter. It appears that the nature of the difference in the distortion might be attributed to those features which have to do with integration and differentiation with respect to distance since it is the placement of stimulus points in the horizontal plane that appears most different between the two figures.

The comparison of data from the two groups of subjects in the Fall and Spring was accomplished by computing correlation coefficients both on group data and also for individual subjects who were present on both testing occasions. These correlations appear in Tables 17 and 18 for dissimilarity judgments and Table 19 and Table 20 for response latency.

In general, the group correlations in Table 17 indicate that both graduate and undergraduate students respond considerably alike from Fall to Spring. As can be seen from an examination of the correlations within the body of the table, however, there are discrepencies when the correlation is broken down by number of transformations. For example, in the combined stimulus set, although the overall correlation between the first trial in the Fall and the first trial in the Spring is .62, the actual

Table 14

Intertrial Correlations for Response and Response Latency (RL) on Combined Set (Graduates and Undergraduates) Spring Testing

Trial	Response		Response Latency	
Trial	Graduates	Undergraduates	Graduates	Undergraduates
1-2	.81	.88	.29	.38
1-3	.84	.93	.20	-.21
1-4	.82	.91	.13	-.06
1-5	.81	.93	.36	-.06
2-3	.96	.94	.43	.47
2-4	.92	.93	.36	.50
2-5	.90	.91	.50	.23
3-4	.89	.97	.39	.36
3-5	.88	.98	.36	.57
4-5	.94	.96	.51	.29

Table 15

Mean Dissimilarity Rating and Standard Deviation
for Each Stimulus Pair in the Combined Set on Last Trial
(Spring Testing)

Category	Stimulus Pair	Graduates		Undergraduates	
		$\bar{X}$	s	$\bar{X}$	s
x	12	4.80	3.56	5.88	3.52
	34	3.00	1.22	3.50	2.14
	56	2.80	2.49	2.62	2.77
	78	3.00	1.22	3.71	2.93
y	51	6.60	3.05	6.25	3.45
	62	7.40	3.29	4.75	2.60
	73	2.40	1.14	6.62	4.17
	84	2.40	1.14	6.88	4.36
z	14	7.60	2.70	5.38	3.02
	23	6.40	3.65	4.75	1.98
	58	5.60	3.65	4.50	2.51
	67	5.80	3.96	3.62	2.56
xy	16	6.80	2.17	7.75	2.49
	25	7.80	2.49	6.38	2.67
	38	5.20	2.17	8.38	2.06
	47	4.00	2.34	8.50	2.67
xz	31	8.40	2.88	6.12	2.47
	42	7.60	3.78	5.50	2.67
	75	6.20	3.56	3.75	2.76
	86	8.20	3.83	4.62	2.62
yz	18	9.40	1.52	7.38	2.62
	27	6.20	4.02	7.00	2.83
	36	8.20	2.95	7.75	2.60
	45	8.80	2.95	8.88	2.53
xyz	53	8.25	3.10	8.00	2.51
	64	8.60	2.79	9.25	2.19
	71	8.80	1.30	7.38	3.02
	82	8.60	1.95	8.12	2.59

Table 16

Mean Response Latency and Standard Deviation for Each Stimulus Pair in the Combined Set on Last Trial (Spring Testing)

Category	Stimulus Pair	Graduates $\bar{X}$	Graduates s	Undergraduates $\bar{X}$	Undergraduates s
x	12	4.68	1.77	3.46	1.09
	34	3.14	1.35	2.61	1.12
	56	6.27	4.00	3.56	1.57
	78	5.79	2.22	8.17	10.69
y	51	5.34	2.85	4.55	2.51
	62	5.23	3.24	3.97	.97
	73	4.08	3.06	4.33	3.12
	84	3.94	2.14	3.52	3.34
z	14	3.12	1.70	3.30	1.35
	23	2.76	.54	2.62	.89
	58	5.16	2.99	4.18	3.15
	67	7.50	2.52	3.40	1.41
xy	16	5.64	2.03	2.63	.55
	25	5.85	1.38	3.70	1.25
	38	7.34	4.76	3.97	1.48
	47	6.63	6.69	2.78	.96
xz	31	3.75	.81	4.00	1.54
	42	2.88	.67	3.41	1.92
	75	7.66	5.06	3.21	1.39
	86	6.42	4.94	7.23	1.34
yz	18	3.94	.65	3.49	1.87
	27	4.44	1.48	3.63	1.74
	36	5.50	2.87	4.09	3.41
	45	4.64	3.02	3.80	4.40
xyz	53	5.04	1.98	5.84	6.16
	64	6.10	6.51	4.85	6.91
	71	4.53	1.59	3.85	1.79
	82	5.86	1.71	4.27	2.18

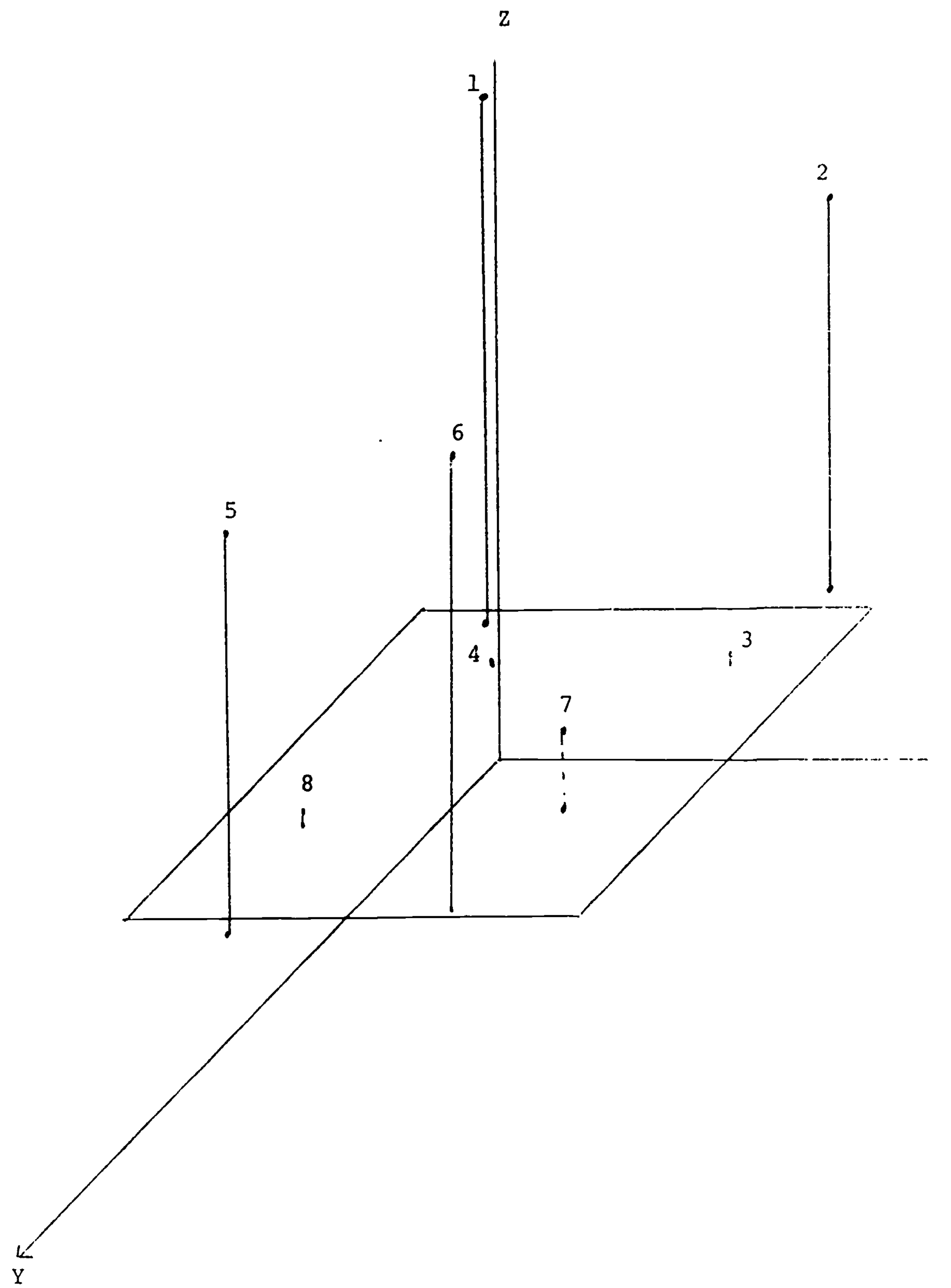


Fig. 3. Graduate Students (Spring Testing), Metric Euclidean, Combined Set.

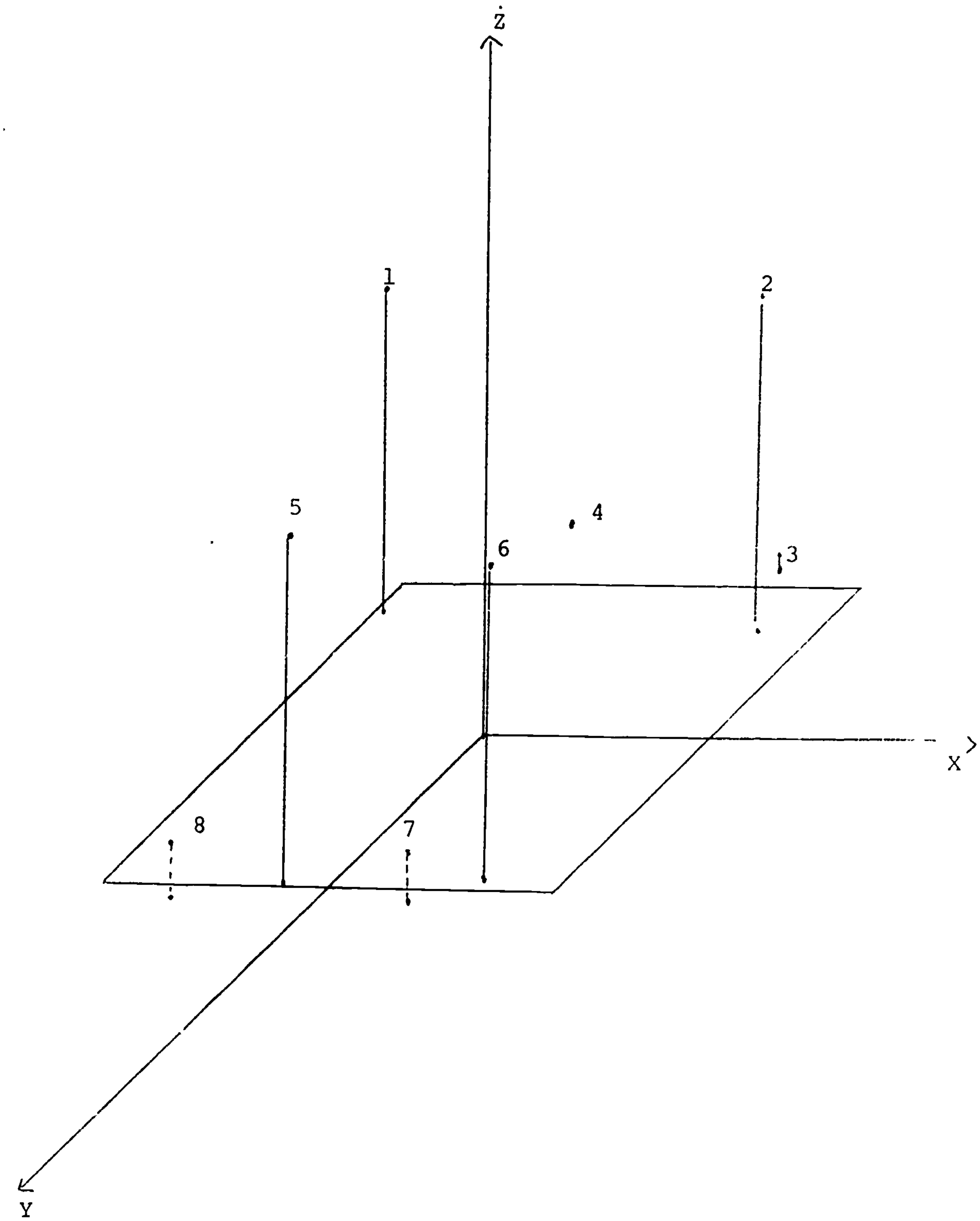


Fig. 4. Undergraduates (Spring Testing), Metric Euclidean, Combined Set.

Table 17

Group Correlations (Fall to Spring) on Dissimilarity Judgments for each Stimulus Set[a]

Combined Set

	Graduates				Undergraduates			
Trials	E	D	G	All	E	D	G	All
1f-1S	.38	.26	.65	.62	.92	.88	.67	.91
2f-2S	.58	.55	-.22	.78	.67	.93	.87	.87
3f-3S	.48	.60	.73	.76	.86	.87	.91	.92
4f-4S	.72	.57	.61	.77	.77	.84	.94	.89
5f-5S	.71	.22	.55	.71	.89	.97	.82	.96

Abstract Set
Graduates

	αβγ				mxt				gm-cm			
	E	D	G	All	E	D	G	All	E	D	G	All
1f-1S	.67	.57	.58	.84	.95	.72	.99	.94	.27	.71	.10	.85
2f-2S	.17	.51	.99	.85	.97	.72	.99	.93	.58	.85	.89	.91
3f-3S	.53	.85	.97	.88	.92	.61	.95	.91	.65	.90	.60	.94

Abstract Set
Undergraduates

	αβγ				mxt				gm-cm			
	E	D	G	All	E	D	G	All	E	D	G	All
1f-1S	.92	.20	.73	.82	.96	.80	.39	.89	.70	.25	.64	.89
2f-2S	.86	.34	.82	.85	.93	.70	.94	.93	.85	.60	-.90	.96
3f-3S	.88	.30	.91	.88	.92	.73	.72	.89	.76	.65	.49	.94

[a]f = fall, s = spring

E = edge, D = diagonal, G = grand diagonal

All = all stimulus pairs

Table 18

Individual Subject Correlations on Last Trial Dissimilarity Judgments (Fall to Spring) for each Stimulus Set[a]

Combined Set

Graduates					Undergraduates				
Subject	E	D	G	All	Subject	E	D	G	All
1	.59	.33	-.80	.38	1	.80	.43	.00	.80
2	.20	-.21	-.33	.12	2	.20	-.02	.00	.28
3	-.25	.45	-.50	.57	3	.28	-.11	-.33	.25
4	-.33	.39	.00	.35	4	.97	.70	1.00	.86
5	.99	.98	.98	.98	5	.85	.77	.98	.81
					6	.94	.79	.87	.89
					7	.52	-.18	.00	.22
					8	.94	.91	.89	.94

Abstract Set
Graduates

	αβγ				mxt				gm-cm			
Subject	E	D	G	All	E	D	G	All	E	D	G	All
1	.08	-.21	.00	.47	.97	.35	.00	.87	.92	.00	.00	.92
2	.90	.92	.99	.87	.87	.69	.98	.82	.00	.83	.16	.57
3	.86	.66	1.0	.75	.51	.84	.86	.03	.81	.94	.00	.90
4	.69	.86	-.57	.77	.12	.67	1.0	.70	-1.0	-.16	.94	.76
5	.88	.31	.00	.83	.70	.51	-.58	.84	.97	.97	.98	.96

Abstract Set
Undergraduates

	αβγ				mxt				gm-cm			
Subject	E	D	G	All	E	D	G	All	E	D	G	All
1	.27	.17	-.31	.46	.37	.20	-1.0	.33	.44	.47	-.40	.53
2	.54	-.27	.87	.31	.93	.26	-.97	.41	.00	-.16	.52	.05
3	.94	.96	1.0	.98	0.00	.99	1.0	.97	.00	.40	.00	.90
4	.62	.52	.00	.80	.93	.63	0.0	.87	.86	.89	.58	.91
5	.24	-.21	.38	.38	.58	.54	1.0	.62	.62	.84	.89	.75
6	-.16	.29	-.58	.31	.94	-.05	.13	.79	.00	.39	.00	.80
7	.92	.88	1.0	.92	.97	.74	.98	.92	.30	.00	.00	.69
8	.86	.09	0.0	.78	.82	-.10	-.33	.83	.68	.64	-1.0	.80

[a] E = edge, D = diagonal, G = grand diagonal, All = all stimulus pairs.

correlation on edge pairs is only .38. Similarily, for the fifth trial in the combined stimulus set in the Fall and the Spring, the overall correlation is .71 while the correlation for diagonals is only .22. If anything, the correlations for transformations are more consistent with overall correlations for undergraduates than they are for graduate students between the two testing occasions.

Moving on to an examination of the individual subject correlations, we find once again more stability for undergraduates than for graduates although now it is apparent that for each group of subjects there is considerable variability as regards how stable responses are from one occasion to the next. Subject number 2, for example, in the graduate group, manifested a correlation of only .12 for all stimulus pairs. This can be compared with subject number 5 in the same group who responded so much alike from Fall to Spring that his correlation averaged .98 for all transformations on all stimulus pairs combined. As was the case with the individual subject analysis in chapter 6, these correlations suggest that utilization of group data on judgment tasks such as have been employed in the present series of experiments, definitely maks individual differences that may carry substantial information about the nature of the knowledge structures which are represented in the distinctive features of alpha-numeric stimulus sets.

The correlations for response latency in Tables 19 and 20 are less informative than the corresponding correlates of dissimilarity judgments. Most correlations are low and in some cases they are negative. The only conclusion that seems warrented at the present time is that the component processes which contribute to the latency of Ss judgment are not well described in our data. Additional tasks such as the one employed in experiment 1 of chapter 3 are necessary if we are to adequately understand the nature of Ss performance.

We conclude from an examination of the data in our longitudinal study that, while group data reflect stability over time both for graduate and undergraduate students, individual subjects change considerably from one testing occasion to the next. This change within subjects is important since it may well reflect an increasing or decreasing capability in the knowledge underlying an individual's processing of the stimulus pairs. Only by indexing subjects with regards to their problem solving proficiency, for example, might we be able to infer whether the change in knowledge structure reflected in individual correlations represents a change in their subject matter competence. The fact that the

Table 19

Group Correlations (Fall to Spring) on Response Latency for each Stimulus Set[a]

Combined Set

	Graduates				Undergraduates			
Trials	E	D	G	All	E	D	G	All
1f-1S	.56	.40	-.23	.33	.37	.54	.07	.45
2f-2S	.69	.26	.68	.40	.66	.39	.96	.28
3f-3S	.14	.33	.71	.15	.20	.19	.41	.16
4f-4S	.24	-.04	.17	.15	-.09	-.10	-.05	.15
5f-5S	.26	-.40	.67	.09	.35	-.11	-.29	.09

Abstract Set
Graduates

	αβγ				mxt				gm-cm			
	E	D	G	All	E	D	G	All	E	D	G	All
1f-1S	-.14	-.19	.25	.26	.19	.45	.87	.38	.09	-.14	.28	.26
2f-2S	.67	.27	-.30	.49	.61	.77	.96	.67	.35	.21	-.47	.27
3f-3S	.64	.01	-.32	-.30	.42	.32	-.13	.27	.59	.01	-.55	-.10

Abstract Set
Undergraduates

	αβγ				mxt				gm-cm			
	E	D	G	All	E	D	G	All	E	D	G	All
1f-1S	.08	.04	-.01	.17	.08	-.12	.15	.17	-.12	.12	.08	.14
2f-2S	.40	.41	-.61	.49	.69	.60	.58	.68	.43	.61	-.45	.72
3f-3S	.37	-.16	-.85	-.06	.39	-.13	.47	.19	.17	-.26	.37	.09

[a] f = Fall, S = Spring

E = edge, D = diagonal, G = grand diagonal,
All = all stimulus pairs

Table 20

Individual Subject Correlations on Last Trial Response Latency (Fall to Spring) for each Stimulus Set[a]

Combined Set

Graduates					Undergraduates				
Subject	E	D	G	All	Subject	E	D	G	All
1	.21	-.60	.51	.03	1	.56	-.26	.67	-.00
2	-.02	.70	.04	.24	2	.43	-.04	-.60	.25
3	-.58	.38	1.00	-.23	3	.58	.57	.37	.52
4	-.17	.32	.78	.25	4	.33	.59	.81	.53
5	-.41	-.24	-.45	-.14	5	.21	.46	-.77	.19
					6	-.08	.39	-.33	.05
					7	.60	-.10	-.81	.14
					8	.36	.26	.97	.36

Abstract Set
Graduates

	αβγ				mxt				gm-cm			
Subject	E	D	G	All	E	D	G	All	E	D	G	All
1	.14	-.51	-.83	.01	.15	.73	-.24	.39	-.23	.37	-.50	.04
2	.73	.27	.58	.49	.25	.49	.78	.56	.74	-.03	.71	.36
3	.11	.28	-.42	.07	-.22	.19	-.16	.03	.42	.70	.48	.55
4	.56	.28	-.21	.23	-.03	-.05	.49	.05	.55	-.08	-.99	.18
5	-1.00	-.18	0.0	-.18	.22	-.39	.38	.06	.54	.62	.86	.41

Abstract Set
Undergraduates

	αβγ				mxt				gm-cm			
Subject	E	D	G	All	E	D	G	All	E	D	G	All
1	-.22	-.46	.00	-.31	.17	-.07	.97	.12	.01	.13	.31	.01
2	.17	-.06	-.86	-.04	.63	-.08	-.63	.26	.12	-.32	-.41	-.08
3	-.04	.21	-.58	.15	.25	.30	-.93	.19	.49	-.06	-.65	.16
4	-.18	-.25	-.54	-.22	.49	-.23	-.71	-.01	.43	.24	-.58	.25
5	.07	.24	-.26	.03	.25	.49	.16	.30	-.14	.05	.16	-.06
6	-.21	.49	-.83	.09	.38	-.03	-.99	.27	.01	-.07	-.03	.30
7	.17	-.48	-.66	-.04	.10	-.29	.61	-.07	-.08	-.02	.88	.30
8	.40	.10	.84	.26	.58	.17	-.02	.35	.73	.40	-.53	.32

[a] E = edge, D = diagonal, G = grand diagonal, All = all stimulus pairs.

correlations do change for individual subjects is encouraging, however, since it is reasonable to assume that the knowledge structures we have presumed to tap with our procedures must change with training such as is received in the ordinary course of educational experience. Were the correlations to remain highly stable over 9 months time, one might be much less optimistic about the opportunities for relating structures of knowledge to problem solving performance.

Chapter 8

Concluding Remarks

The series of experiments presented in this report was designed to attack the problem of representing the psychological structure of knowledge in a non-trivial domain of semantic content. Because of the difficulties inherent in any semantic analysis, the domain was chosen from a highly restricted area of semantic reference. The use of a technical language, particularly one that employs mathematics, makes the task of constructing a semantic model considerably easier than would be the case for most kinds of stimulus features.

On the basis of our data we conclude that similarity judgments can be used to index the way in which individuals perceive complex alpha-numeric displays. In particular, these judgments reflect the structural properties that are built in to displays and which allow them to be related to one another in regular ways.

The fact that group data for varying levels of subject matter mastery are highly similar can be regarded as a strength rather than a weakness in the present analysis, for it is in individual performance that true knowledge is revealed. The most powerful finding of our research is very likely the fact that similarity in group data masks differences between individuals which are a function of their level of subject matter mastery.

When the stimulus field is a largely meaningless (but nevertheless regular) display of symbols, background and training make the responses of different subjects much alike. When the stimulus field can be interpreted on the basis of a common educational experience, however, individual subjects differ considerably with respect to the way they respond to the various features which underlie stimulus complexity. One may suppose that this variability is indicative of variability in the structures which support and guide problem solving behavior.

Because mastery level can also be regarded as a function of time spent on the judgment task, we speculate that the judgment task reflects an analysis of the stimulus field for conceptual categories (such as force and acceleration) which have been used to mark or classify the apparent features in some educational context.

The fact that these categories are consistent with the more primitive analysis performed by the psychology students is interesting and suggests that the language of the subject matter is based upon abstractions of considerable generality.

The research reported here also gives evidence to the fact that one can establish psychophysical functions for the kinds of stimulus complexes that are used to carry meaning in our educational experience. Knowledge of the nature of these functions can inform us about the way in which individuals learn and think about the content of that experience. In particular, these functions can be regarded as evidence for coding schemes which permit the storage, retrieval, and ultimately the generation of large amounts of technical information. A scheme such as the one we have proposed would permit this kind of activity even though its users were largely unaware of its existence.

The virtue of the type of task we have employed is that it does not require the skills ordinarily associated with work samples of subject matter behavior. This means that with further research, such tasks might be used not only to assess subject matter mastery, but more importantly, to give evidence of deficiencies or inadequacies in the process of acquisition. It is our belief that the knowledge structures revealed on tasks employed in this research can serve to index the adequacy with which an individual understands the knowledge he calls forth in demonstrating and utilizing his subject matter expertise.