

Development of Stimulus Classes as a Function of
Training Procedures and Diversity of Exemplars

by

Donald J. Salmon

A thesis submitted to the Faculty of Graduate Studies
in partial fulfilment of the requirements for the Degree of
Doctor of Philosophy

Department of Psychology

University of Manitoba

May 1, 1991



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-77036-8

Canada

DEVELOPMENT OF STIMULUS CLASSES AS A FUNCTION OF
TRAINING PROCEDURES AND DIVERSITY OF EXEMPLARS

BY

DONALD J. SALMON

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY

© 1991

Permission has been granted to the LIBRARY OF THE UNIVER-
SITY OF MANITOBA to lend or sell copies of this thesis, to
the NATIONAL LIBRARY OF CANADA to microfilm this
thesis and to lend or sell copies of the film, and UNIVERSITY
MICROFILMS to publish an abstract of this thesis.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

Acknowledgements

I would like to thank the members of my committee, Dr. G. Martin, Dr. J. Whiteley, Dr. J. Walker and Dr. T. Stokes (University of South Florida) for their assistance with my dissertation. I am particularly grateful to my advisor, Dr. J. Pear, for his support through-out my graduate career.

I would also like to express my appreciation to Cyndy Erickson, Cindy Cain, Bea Lough, Daryl Owens, Beth Clark, Audrey Langlois, Denise Whynot and the other members of the support staff within the Department of Psychology and the Psychological Service Center who provided much needed assistance throughout the many years of my graduate training.

Also, I would like to thank Dr. Carl Stephens, Valdine Huyghebaert, Tsippi Nahir, Marcie Desrochers, Rosemarie Hrydowy, Ron Kibbins and the St. Amant Daycare staff, for their support and assistance during the completion of this research.

I would also like to express my gratitude to my parents, Jim and Phyllis Salmon, who provided many years of emotional and financial support.

I would like to thank my son, Bruce, who provided much needed humor and moments of distraction with our conversations on the state of the music world. He has grown up to be such a pleasant young man during the course of my graduate career.

I would like to express my sincere appreciation to my beautiful wife, Laurel Lee. Through-out the years she has been my biggest fan and handled my mood changes with wit and charm. She continues to be my lover, friend and companion and it is to her that I dedicate this dissertation.

Table of Contents

Abstract	1
Introduction	3
Statement of the Problem	16
Experiment 1	18
Method	18
Subjects	18
Setting and Apparatus	19
Experimental Design	22
Preliminary Procedures	22
Preliminary Assessment	23
Imitation Assessment	24
Instruction-Following Assessment and Token Training	24
Overview of Training and Testing Procedures	25
Laboratory Test	26
Modification of the Laboratory Test	29
Training Procedures	30
Training One Exemplar at a Time	31
Alternating Procedure	34
Post-Training Test for Learning	36
Generalization Test	38
Dependent Variables	38
Interobserver Reliability	39
Results	42
Discussion	56

Experiment 2	60
Method	60
Subjects	60
Setting and Apparatus	61
Training and Testing Stimuli	61
Experimental Design	66
Preliminary Procedures	67
Overview of Training and Testing Procedures	67
Dependent Variables	68
Interobserver Reliability	68
Procedural Reliability	68
Results	70
General Discussion	83
References	88
Appendix A	92
Appendix B	93
Appendix C	102
Appendix D	104
Appendix E	107
Appendix F	108
APPENDIX G	109
Appendix H	114
Appendix I	115
Appendix J	116

Abstract

Two experiments were conducted in order to study generalization within stimulus classes based on common physical characteristics. In Experiment 1, using a replication design, the effects of two different training procedures were examined with three developmentally delayed children. In one condition six stimulus classes were trained, using one member from each stimulus class. Once training had been completed, generalization was assessed to seven other members from each stimulus class trained. Following the generalization test, the stimulus classes were retrained using a second member from each of the six stimulus classes. These were the stimulus members judged least similar to the original training stimuli. Generalization was assessed to the remaining stimulus class members. In a second condition six stimulus classes were trained by alternating the two stimulus members least similar to one another. Following training, generalization was assessed to untrained stimulus class members. The results from this experiment indicated that both procedures produced the same amount of generalization, although the training of two exemplars near the same point in time may be more cost-effective when training a large number of stimulus classes.

Experiment 2 examined the issue of similarity and diversity of exemplars. One developmentally delayed and two normal children were taught to name stimulus classes within a replication design. In one condition, six stimulus classes were trained using two similar members from each stimulus class. Generalization was assessed to stimulus class members rated as being similar and different to the training members. In a second condition, six stimulus classes were trained using two stimulus members from each class that had been rated as different from one another. Generalization was again assessed to similar and different untrained stimulus class members. The results indicated that the training of two diverse stimulus members produced high amounts of generalization to both similar and different untrained stimulus members. The training of two similar stimulus members proved effective in promoting generalization to other similar stimulus members, but not to stimulus members which reflected greater diversity.

The results from both experiments were also analyzed on a post hoc basis to determine if they would add support to the "best exemplar theory" of categorization (Mervis & Pani, 1980). The results of this analysis did not support this theory. Recommendations for further research are suggested.

Introduction

Generalization is an important consideration for anyone concerned with affecting any kind of behavioral change. The present research is concerned with stimulus generalization within stimulus classes in the area of language development. Stimulus generalization is often said to have occurred when a response becomes more probable in the presence of a novel or new stimulus or situation after having been reinforced in the presence of another stimulus or situation (Martin & Pear, 1988). However, the strength of the behavior with the new or novel stimulus is less than that found with the original training stimulus. Furthermore, this decrease in strength is related to the difference between the training stimulus and the new or novel stimulus (Guttman & Kalish, 1956).

Stokes and Baer (1977), in their seminal article on generalization, defined it in a somewhat different manner:

"generalization will be considered to be the occurrence under different nontraining conditions (i.e., across subjects, setting, people, behavior, and/or time) without the scheduling of the same events in those conditions as had been scheduled in the training condition. Thus, generalization may be claimed when no extra training manipulations are needed for extra

training changes; or may be claimed when some extra manipulations are necessary but their cost or extent is clearly less than that of the direct intervention (p. 350)."

It would seem that these authors approach generalization from more of a cost-benefit analysis. In addition, they would appear to define it in a much broader sense in that they are discussing response generalization, stimulus generalization, and behavioral maintenance (Martin & Pear, 1988). The present research, when referring to aspects of the Stokes and Baer article, will be concerned with its suggestions for the promotion of stimulus generalization.

Within the area of language development and developmental disabilities, a question which has received some attention from researchers concerns the amount of generalization found across stimulus modalities (e.g. objects, picture cards) after training in one modality. Previous research findings have shown that a naming response taught using picture cards or photographs of objects will not automatically generalize to the corresponding real object when they are encountered by an individual in the natural environment. For example, Welch and Pear (1980) taught naming responses to retarded children in a classroom using either the real object, a photograph of the object, or

a picture of the object. Generalization to real objects was assessed in the children's natural environment. For three of the four children tested, generalization was found to be superior after training with objects. All three training modalities were found to be equally effective with the fourth child. These results support the use of real objects if one wants to ensure generalization to real objects when encountered in the natural environment.

In a similar vein, Handleman, Powers, and Harris (1984) taught either objects or pictorial representations of objects to three retarded children and tested for generalization across modalities. For two of the three children, generalization was low while the results with the third child were considerably better. Thus it would seem that the ability to generalize from picture cards to objects and vice versa is not an automatic process and needs to be programmed.

Related to the question of generalization from objects to picture cards and vice versa is the question of the amount of generalization found within a stimulus class after training with one member of that class. As defined by Martin and Pear (1988), a stimulus class "is a set of stimuli all of which have some characteristic in common" (p. 182). For example, all objects which we commonly label with

the word "cup" share a similar shape and function. When an individual can correctly label all "cups" and discriminate between "cups" and other objects he or she is showing conceptual behavior (Keller & Schonfeld, 1959). Spradlin (1982) has defined a stimulus class in a somewhat different manner to refer to a number of stimuli which all control the same behavior. This definition is much broader than that used by Martin and Pear since presumably stimuli which do not share common characteristics could be potentially members of the same stimulus class. However, the present research is concerned with the development of stimulus classes that share certain characteristics and will thus rely on the Martin and Pear definition. Specifically, the term "stimulus class" as used in the present research will refer to a number of objects which have certain characteristics in common and are labelled with the same name or verbal response (e.g., cup, kettle, bracelet).

In an extension of the work of Welch and Pear (1980), Salmon, Pear, and Kuhn (1986) trained retarded children in a classroom to name either picture cards or real objects that were selected on the basis of matching the picture cards. Once training had been completed, generalization was assessed in the child's natural environment using the training object or the object that matched the picture card (if picture cards had been trained) plus four other objects

belonging to the same stimulus class as the training stimulus. The results replicated those of Welch and Pear (1980) in that generalization to the training objects was greater after training with the real objects than after training with pictures of the objects. Furthermore, generalization to the other stimulus class members after training with objects was superior to generalization found after training with picture cards. However, regardless of whether training was conducted using objects or picture cards, generalization was far from ideal for most stimulus classes. It should be noted, however, that after training with real objects generalization to all stimulus class members was at or near optimal levels for a few stimulus classes.

In a second experiment, Salmon et al. (1986) attempted to increase the amount of generalization by retraining half of the naming responses which had initially shown little generalization by alternating either the original training object or the object that matched the original training picture (in those cases where picture cards had been the training stimulus) with the object least similar to the training object. The other half were simply retrained using the original training stimulus or the object that had matched the picture card. For the three children who participated in this research, the alternating procedure

resulted in significant increases in generalization whereas there was only modest or no gains in generalization after simply retraining the naming responses using the original training stimulus. This method of promoting generalization is an example of training sufficient exemplars (Stokes and Baer, 1977).

Similar findings have been noted by Anderson and Spradlin (1980) who trained a retarded child to name different objects and tested for generalization to other objects belonging to the same stimulus classes as the training objects. For four of the six object names trained, the use of only one exemplar was insufficient to produce generalization to other members of the same stimulus class. However, the training of more exemplars produced systematic increases in the amount of generalization obtained. For two of the naming responses the training of three exemplars was sufficient to produce perfect generalization. For one, it was necessary to train four exemplars; and for the remaining object, five exemplars had to be trained before perfect generalization occurred.

The results of the study by Salmon et al. (1986) and the study by Anderson and Spradlin (1980) clearly indicate that if one wishes to increase generalization within a stimulus class, the training of more than one exemplar is an

important requirement, although there were exceptions to this general rule in both studies. While both the above studies were similar in terms of their goals, there was some important procedural differences. In the Salmon et al. study, during training the authors concurrently trained two stimulus classes, alternating two exemplars from each stimulus class, and tested for generalization. On the other hand, Anderson and Spradlin, using either a concurrent or a serial training procedure, trained one exemplar at a time and tested for generalization. If insufficient generalization was noted a second exemplar was trained, and so on until generalization was at an optimal level.

The procedural differences between these two studies are not uncommon in the "sufficient exemplar" literature. In their seminal article, Stokes and Baer (1977) suggest that if the training of one exemplar is insufficient in terms of producing adequate generalization then a second and a third exemplar, etc., should be trained until generalization is at the desired level. This suggestion of training one exemplar at a time is often found in the behavioral literature. For example, Parsonson and Baer (1978) trained tool improvisations with preschool children by training one exemplar of a particular tool (e.g., hammer) and noted whether generalization occurred. The authors

introduced a new tool each session and noted the effect on generalization to untrained tools.

Research which is similar to that of Salmon et al. (1986) is also often found in the sufficient exemplar literature. Stokes, Baer, and Jackson (1981) found little generalization of a greeting response after training with one experimenter. However, training of this greeting response concurrently with two experimenters resulted in generalization of the greeting response to a number of experimenters.

Thus, while similar, the above mentioned studies have some important procedural differences. However, it may very well be that the Salmon et al. procedure represents a more fruitful means of promoting generalization in that the alternating of two members of the same stimulus class at about the same point in time may increase the probability that the behavior is more quickly brought under appropriate stimulus control. It would seem likely that by systematically varying members of a stimulus class near the same point in time, thus exposing an individual to the relevant and irrelevant physical characteristics of the stimulus class along with the verbal label for the class, one should be successful in teaching an individual which

properties are relevant in terms of being able to correctly name or identify a particular stimulus class.

Also very much related to the issue of generalization is the question of how many exemplars are sufficient in terms of producing the desired amount of generalization. At present no succinct answer can be forthcoming, although Stokes and Baer (1977) suggest that the type of task and the subject's prior history are important determinates. In further commenting on this issue they note that:

Diversity of exemplars seems to be the rule to follow in pursuit of the maximum generalization. Sufficient diversity to reflect the dimensions of the desired generalization is a useful tactic. However diversity may also be our greatest enemy: too much diversity of exemplars and not enough (sufficient) exemplars of similar responses may make potential gains disproportional to the investment of training effort. The optimal combination of sufficient exemplars and sufficient diversity to yield the most valuable generalization is critically in need of analysis. (Stokes and Baer, 1977, p.357)

In a recent article Stokes and Osnes (1989) have attempted to refine the classification system proposed by

Stokes and Baer (1977). They proposed three areas of general principles: a) Exploit current functional contingencies; b) Train diversely; and c) Incorporate functional mediators. Within each area four programming tactics are described. Of particular concern to the present research is the general principal of "Train Diversely". Within this area, one of the four programming tactics they describe is the use of sufficient stimulus exemplars. While the authors suggest that only a small number of stimulus exemplars is needed to promote generalization, they do not specifically comment on how similar/diverse exemplars need to be for the programming of generalization.

The Salmon et al. study used only two exemplars (the most diverse) and found overall increases in generalization, although for some stimulus classes the alternating procedure did not produce increases in generalization to optimal levels. In the Anderson and Spradlin study (1980) one exemplar had to be trained in two cases only, while with the remaining objects a range of from 3-5 exemplars had to be trained before generalization was found to be optimal to all testing stimuli. While in the Salmon et al. study the authors alternated the two least similar members of a stimulus class, in the Anderson and Spradlin study the authors did not and seemed to select the order of presenting exemplars on a purely random basis.

It may very well be that in both above mentioned studies, where only one exemplar had to be trained the testing stimuli may have been very similar to the exemplar trained, thus reflecting little diversity. In other cases where the alternating procedure was used and revealed little generalization (Salmon et al., 1986) or when four or five exemplars had to be trained before generalization was found to be sufficient (Anderson & Spradlin, 1980), the initial stimuli used may have reflected too much diversity and little similarity.

Within the behavioral literature the question of diversity and number of exemplars has gone largely unanswered. However, within cognitive psychology these questions have received some research attention especially as they pertain to increasing generalization within a stimulus class. From a cognitive perspective the above problem is one of categorization. Categorization occurs when two or more objects are treated as being equivalent (Mervis & Pani, 1980). Within the cognitive perspective one of the best known theories of categorization is that of the best-example theory (Rosch, 1973). According to this theory the best examples of a particular category are those which have a number of attributes which are possessed by other members of the same category while at the same time having

few of the attributes possessed by other categories (Mervis & Pani, 1980).

Typically in this type of research, adults are asked to rate a number of objects or photographs of objects, on a 7 point scale, as to whether the stimulus is a good example of the particular category, a poor example of the category, or falls somewhere in between (Hupp, 1980). The stimuli of a category are then classified in the following manner: good 1-3.19, medium 3.2-4.19, and poor 4.2-7. Using the above method, Mervis and Pani (1980) trained kindergarten children to name 6 "nonsense" categories using "nonsense" syllables. In one condition subjects were trained using one good example from each category and tested receptively to see how well the response generalized to three other objects belonging to the same category. In a second condition one poor exemplar from each category was trained and generalization was again assessed to other members of the categories. The results revealed that generalization to other category members was significantly better after training with one good exemplar than with one poor exemplar.

In a similar study severely retarded children were trained to name objects using three different conditions (Hupp & Mervis, 1982). Again, adults initially rated a number of pictures of objects as to whether they were good,

poor, or moderate examples of a particular category. Each category contained 26 pictures of objects, with two categories being trained in each of three conditions. In one condition one good exemplar (GE1) of each category was trained; in the second condition three good exemplars (GE3) of each category were trained; and in the third condition one good, one moderate, and one poor exemplar were trained. The authors examined both expressive and receptive language and report that for expressive language, training with one or three good exemplars produced more generalization than training with one good, one moderate, and one poor exemplar of a category. They also noted that training with three good exemplars resulted in generalization to a greater range of exemplars (good, moderate, and poor) than did training with one good exemplar. Unfortunately, the authors present no data, statistical or otherwise, to support this claim. Thus their conclusions cannot be properly evaluated.

When the children's receptive language abilities were statistically analyzed the authors, in general, found results similar to those found with expressive language. Generalization to all exemplars was statistically significant for the GE3 condition, although the overall mean percentage of correct responses to all untrained photographs was only 50.8%. In the GE1 condition only those responses to exemplars which had been classified as good reached

statistical significance. The mean percentage of correct responses to these exemplars was 50%. In the condition in which a good, moderate, and poor exemplar were trained, the results failed to reach statistical significance.

In commenting on their findings the authors suggest that what is important is not so much the number of exemplars trained (cf. Stokes and Baer), but whether these exemplars are good examples of a given category. This would seem to imply that similarity of (good) exemplars rather than diversity is crucial in terms of training stimulus classes. However, in the Salmon et al. (1986) study it was the two most extreme or diverse exemplars which were found to be effective in terms of increasing generalization.

Statement of the Problem

The above review of the literature has identified two unresolved issues related to the development of stimulus classes specifically, and to the generalization literature in general. The first issue concerns the means by which a stimulus class is trained; i.e., through the training of one exemplar of a stimulus class at a time or by alternating two exemplars of a stimulus class at about the same point in time. The second issue relates to the question of diversity of exemplars. Is the more efficient means of promoting generalization to train two exemplars that are similar to

one another or two exemplars that reflect a wider range of stimulus characteristics. The proposed research attempted to answer these questions. In the first experiment children were trained to name objects using the two exemplars least similar to one another as training stimuli (cf. Salmon et al, 1986). In one condition these two stimuli were trained one at a time with generalization to untrained stimuli being assessed after each had been trained. In the second condition the two trained stimuli were trained at the same time using the alternating procedure (cf. Salmon et al, 1986) and generalization to untrained stimuli was again assessed.

A second experiment examined the question of diversity of exemplars. Using the alternating procedure, two exemplars were trained. In one condition the exemplars used in training were selected to reflect greater similarity and less diversity. In the second condition the exemplars were selected to reflect greater diversity and less similarity. This experiment should shed some light on the issue of similarity and diversity of exemplars which, as Stokes and Baer (1977) have suggested, is critically in need of analysis. The present results should greatly add to our understanding of the most effective means of teaching stimulus classes to normal and developmentally delayed individuals.

Experiment 1

MethodSubjects

Three students from the Brockville public school in Spruce Grove, Alberta, participated in the present research. This school has classes from Grades K to 9, with handicapped children being integrated into the regular school system.

Potential subjects for the research were selected on the basis of interviews with teachers. A letter was sent to the parents of the children explaining the nature of the research and seeking permission to employ their child in the research. A token reinforcement system was used throughout the study because some teachers were concerned about providing extrinsic "rewards" to the children. They saw the token system as being more acceptable in that the primary reinforcers followed the training sessions. This, they believed, more closely resembled the classroom situation where they might provide "treats" to the children after completing a lesson.

Dawn was six years old and diagnosed as being developmentally delayed, especially in the area of speech. Her last formal speech assessment, at the age of 4-2, had found her receptive language equivalent to age 2-2. Her

expressive language abilities were equivalent to age 1-6. She also suffered from a mild to moderate hearing loss in her right ear and as a result wore a hearing aid.

Esther was 5 years old and had been diagnosed as having Mosaic Down's Syndrome. She was described as having a short attention span. A psychological assessment conducted independently at the same time this research was being conducted found her to have an IQ of 61 ± 6 , using the Stanford Binet Test of Intelligence.

Katherine was 8 years old and diagnosed as Autistic. On the Peabody Picture Card test she scored a MA of 2-1. Her expressive language abilities were found to be equivalent to a 2-3 year old.

Setting and Apparatus

Training and testing sessions were conducted in a 3 m x 4 m room in the school. The experimenter and the subject sat opposite one another at a table. A small console was placed on the table which allowed the experimenter to control the length of a trial and intertrial interval. A trial lasted for 8 seconds or until the child emitted a response, whichever came first. This was followed by an intertrial interval of 5 seconds. Training and testing

objects were placed on the floor within easy reach of the experimenter but out of sight from each child.

Nearby a clock timed the length of each session and a camcorder videotaped each session. A tape recorder, on the table, audiotaped all sessions. Data sheets and token reinforcers were placed on the table in front of the experimenter. Each session was 20 minutes in length.

Training and Testing Stimulus

Forty stimulus classes were selected prior to the start of this research. As much as possible, objects were selected on the basis of being commonly found in a person's environment. In addition, ease of presentation was also considered in determining which stimulus classes would be trained. Within each stimulus class, eight objects (stimulus members) were selected for training and testing purposes. Initially one member from each stimulus class was selected by the experimenter and designated the first training member (SM1). Following this, 10 adult raters independently rank ordered the remaining seven members in relation to the first training stimulus member on the basis of similarity (See Appendix A for the rating instructions). Since the most extreme stimulus members were to be trained, if seven of the ten raters agreed on which member was least similar to the first training member, the eight members of

the stimulus class were deemed acceptable for the purposes of the research. However, if fewer than seven of the ten raters agreed on which stimulus member was least similar to the first training member, the experimenter selected new stimulus members for that stimulus class and they were then rated a second time, etc., until the desired level of agreement was reached. In addition, if three or more of the ten raters indicated that a stimulus member did not belong to that stimulus class it was discarded and a new one added. (The final rating for each stimulus class are shown in Appendix B.)

Furthermore the experimenter had 10 raters independently rate members of each stimulus class as to whether they were good, moderate, or poor examples of a stimulus class (category) using the procedures developed by Mervis and Pani (1980). (The instructions for these ratings are in Appendix C; the results of the ratings are in Appendix D.) It should be noted that the objects were selected to conform to the rules laid out in Appendix A. No attempt was made to ensure that an equal number of stimulus members were classified as good, moderate, or poor examples of a category (cf. Mervis & Pani, 1980). The information obtained from the categorization ratings was used post hoc in an attempt to integrate the outcome of this research with the research findings from cognitive psychology.

Experimental Design

Stimulus classes were trained using a design involving replication of two conditions. In one condition one stimulus member at a time was trained (cf. Anderson & Spradlin, 1980). In the other condition two stimulus members were trained at the same time using the alternating procedure (cf. Salmon et al., 1986). In both conditions only two members of each stimulus class were trained. These were the ones determined on the basis of the ratings to be least similar to one another. Within each phase each child was trained to name six stimulus classes. The stimulus classes were randomly assigned to each condition. Due to problems with the amount of generalization found in the first phase with one child, there were only enough stimulus classes to complete one replication of one of the conditions. In addition, for one child, due to time constraints, each condition was presented only once.

Preliminary Procedures

Prior to starting the research, each child proceeded through a number of preliminary steps in order to ascertain his or her suitability for the proposed research. These steps were: (1) Preliminary Assessment; (2) Imitation Assessment; and (3) Instruction Following Assessment. It should be noted that when proceeding through these

preliminary steps and during the experimental phases each subject was worked with on an individual basis.

Preliminary Assessment

The function of the preliminary assessment was to serve as a screening device in order to ensure that the children were suitable candidates for the proposed research. The experimenter presented each member of all 40 stimulus classes, one at a time, followed by the probe: "What's this?". The subject had 8 seconds in which to name the object. This procedure was repeated a second and third time such that each member of all stimulus classes had been presented to a child three times. All correct responses to any of the members of a stimulus class were given primary and social reinforcement. Each member of a stimulus class which the subject failed to correctly name on all three occasions was considered unknown and suitable for training. If a subject correctly named one or more members of a stimulus class on one or more occasions it was considered known for the purposes of this research and unsuitable for training.

After the completion of the preliminary assessments, subjects were selected on the basis of failing to name, at a minimum, 24 stimulus classes. All three children met this minimum criteria. If more than 24 stimulus classes were

unknown, the experimenter randomly selected the 24 stimulus classes to be used in the research.

Imitation Assessment

Immediately after the preliminary assessments were completed an imitation assessment was conducted. The experimenter removed all training and testing stimuli and said to the subject: Say "_____". Each of the naming responses was presented sequentially, three times each. The experimenter noted whether the subject could imitate the response correctly or give a reasonable approximation of the desired response. Responses which the subject failed to imitate correctly were given imitation training until the subject could emit an exact or close approximation of the desired response.

Instruction-Following Assessment and Token Training

Following the imitation baseline each subject was assessed as to her ability to follow some simple instructions (e.g. touch your nose, touch your hair, clap your hands). If a subject was unable to follow some of these instructions, she was trained to emit the responses upon request, through the use of modelling and physical guidance.

The experimenter initially modeled the desired response. If the subject imitated the modeled response she was reinforced. If the subject failed to imitate the modeled response, the experimenter modeled the correct response and at the same time physically guided the subject in making the response. After the behavior was performed at a high rate, the modelling and physical guidance were faded out until the subject was emitting the response on request. All correct responses were given tokens in the form of pennies and social reinforcement in the form of "Good girl". Following the end of each session tokens were exchanged for backup reinforcers. Backup reinforcers were a choice between pieces of "Smarties", chips, "Cheezies", or 1/2 teaspoon of ice cream. However, over the course of assessing and training instruction following, the schedule of reinforcement for the tokens was changed until the subject was reinforced on a variable-ratio (VR) 5 schedule of reinforcement (on the average every fifth correct response was reinforced).

Overview of Training and Testing Procedures

Each phase of the experiment proceeded through the following steps: (1) Laboratory Test; (2) Training; (3) Post-Training Test; and (4) Generalization Test. This sequence differed somewhat for the two experimental conditions. In one condition six stimulus classes were

trained using only the first stimulus class member (SM1). Once all six stimulus classes reached criterion, the Post-Training Test and Generalization Test were conducted. This was followed by retraining the same stimulus class using the stimulus member least similar (SM8) to SM1 as the training stimulus. Once the stimulus class had again reached criterion, a second Post-Training Test and Generalization Test are conducted. In the other condition the six stimulus classes were trained to criterion alternating stimulus class members SM1 and SM8. This was followed by the Post-Training Test and the Generalization Test. Thus training in this condition followed the procedure outlined by Salmon et al., (1986).

Laboratory Test

Once the 24 stimulus classes were selected, six were randomly chosen for each phase of the study for each subject. During the laboratory test all eight members of the six stimulus classes were reassessed in order to ensure that they continued to be unknown. The eight members from each of the stimulus class were randomly presented to each of the subjects, one member at a time, followed by the probe "What's This?", whereupon the subject had 8 seconds to respond with the appropriate name of the object. These probes were interspersed among 14 instructions. Probes and

instructions were interspersed in the following manner (with reinforcement indicated in parenthesis):

- 1) Instruction
- 2) Instruction (Reinf.)
- 3) Instruction
- 4) "What's This?"
- 5) Instruction
- 6) Instruction
- 7) "What's This?"
- 8) Instruction (Reinf.)
- 9) Instruction
- 10) Instruction
- 11) Instruction (Reinf.)
- 12) "What's This?"
- 13) Instruction
- 14) "What's This?"
- 15) Instruction
- 16) "What's This?"
- 17) Instruction (Reinf.)
- 18) Instruction
- 19) "What's This?"
- 20) Instruction

This sequence of instructions and probes was repeated 24 times such that all eight members of the six stimulus classes were presented three times. The order of

presentation of the training items was randomized within each sequence as was the presentation of instructions.

The above procedure is identical to that used by Welch and Pear (1980) and Salmon et al.(1986). The rationale for reinforcing only correct responses on instruction trials is based on research by Bucher (1973) and Martin (1974), who noted that reinforcing only certain instruction trials was sufficient to maintain correct responses to nonreinforced instructions. Thus, presumably, reinforcing only correct responses to certain instructions enabled the experimenter to assess the amount of generalization which occurred without confounding the results with reinforcement. At the same time maintenance of the behavior during Generalization Tests was ensured.

Those members of a stimulus class, which a subject again failed to name on all three occasions on which the above sequence was presented, were considered unknown and suitable for training. If a subject correctly named a member of a stimulus class on one or more trials, a new stimulus class was selected and the Laboratory Test was repeated.

Modification of the Laboratory Test

As a function of the results obtained in the first training condition with Dawn, some procedural changes were made to the Laboratory Test and this training condition was replicated for both Dawn and Esther. Within the Laboratory Test, the instructions were dropped and known stimulus classes were interspersed with unknown stimulus classes. In addition, because the Laboratory Test was found to be rather time consuming, the ratio of known to unknown trials was increased from 14:6 to 10:10. The higher ratio resulted in fewer trials overall since there was a limited number of unknown stimuli. Probes for known and unknown stimulus classes were interspersed in the following manner:

- 1) "What's This?" (Unknown)
- 2) "What's This?" (Known) (Reinf)
- 3) "What's This?" (Unknown)
- 4) "What's This?" (Unknown)
- 5) "What's This?" (Known)
- 6) "What's This?" (Known)
- 7) "What's This?" (Unknown)
- 8) "What's This?" (Known) (Reinf)
- 9) "What's This?" (Unknown)
- 10) "What's This?" (Unknown)
- 11) "What's This?" (Known) (Reinf)
- 12) "What's This?" (Unknown)
- 13) "What's This?" (Known)

- 14) "What's This?" (Unknown)
- 15) "What's This?" (Known)
- 16) "What's This?" (Unknown)
- 17) "What's This?" (Known) (Reinf)
- 18) "What's This?" (Known)
- 19) "What's This?" (Unknown)
- 20) "What's This?" (Known)

Thus 10 unknown probes were interspersed among 10 known probes. The schedule of reinforcement was unchanged and only known probe items received social or token reinforcement.

Training Procedures

Training was conducted four to five days per week. One 20 minute session was run. The last 5 minutes of each session was devoted to cashing in the tokens for back-up reinforcers. A subject proceeded through a series of prompt and probe trials until the stimuli reach criterion and were considered known. During a prompt trial the experimenter named the stimulus (e.g. "What's This?' Kettle'."). The child was required to imitate the prompt. During a probe trial the subject was asked to name the stimulus (e.g. "What's This?"). A concurrent training procedure was used, similar to that employed by Rowan and Pear (1985) and Salmon et al. (1986). The experimenter randomly selected two

stimulus classes for training purposes. While only two stimuli from each stimulus class were trained in both conditions, the manner in which they were introduced to the subjects differed. The two different training procedures are described below.

Training One Exemplar at a Time

In the one-exemplar-at-a-time (OET) procedure the first training members (SM1s) were initially trained according to the following 8 step sequence (See Figure 1):

- 1) In Step 1 the experimenter held up SM1 from the first unknown stimulus class and presented a probe trial. This first trial acted as a check against the stimulus classes being known prior to training. If the subject responded correctly to the training stimulus on this trial and this trial only, it was discarded and a new stimulus class was selected. If the subject made an error (i.e., an incorrect response or failed to respond within 8 seconds) a prompt trial was presented. An error on this trial resulted in it being repeated. A correct response to the prompt resulted in the probe trial being presented again. A correct response to the probe trial in Step 1 resulted in Step 2.

- 2) Step 2 was the same as Step 1 with the second unknown stimulus class.

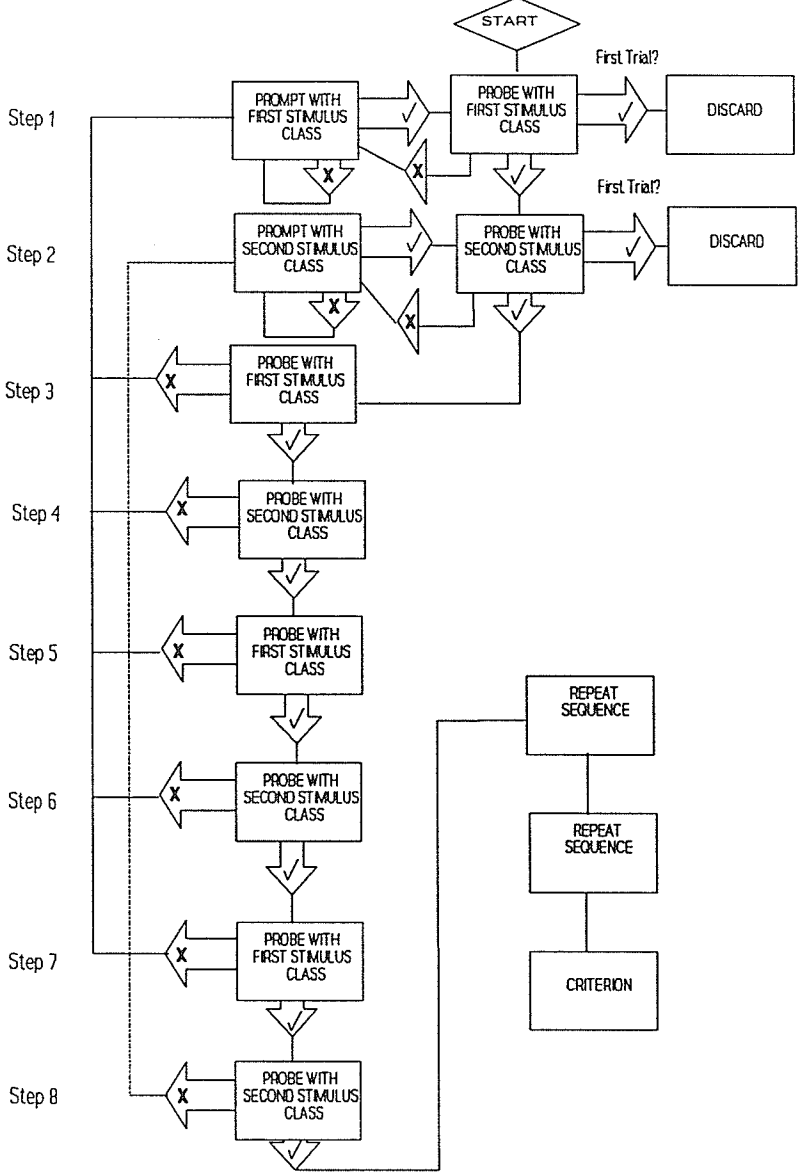


Figure 1 Flow chart of training procedures

3) Step 3 was the same as Step 1, except: (a) the stimulus was not discarded if a correct response occurred on the first probe trial; (b) an error resulted in a return to Step 1.

4) Step 4 was the same as Step 3 with the second unknown stimulus class, except that an error resulted in a return to Step 2.

5) Step 5 was the same as Step 3.

6) Step 6 was the same as Step 4.

7) Step 7 was the same as Step 3.

8) Step 8 was the same as Step 4. A correct response on a probe trial resulted in the experimenter recycling through the training procedure.

Successful completion of the training procedure three times resulted in the stimulus classes being said to have "reached criterion" and considered known. During the last two training sequences a correct response to the first probe trial of Steps 1 and 2 did not result in the stimuli being discarded.

The remaining four stimulus classes were then trained, two at a time, using the SM1's from each of the remaining stimulus classes. When all six stimulus classes had been trained the Post-Training and Generalization Tests were conducted. Once these two tests were completed the stimulus classes were retrained using the second stimulus member (SM8). The training procedure was identical to that used with SM1 with the exception that a correct response to the first probe of Steps 1 and 2 did not result in the stimuli or stimulus classes being discarded. During training all correct prompts were given social reinforcement in the form of "Good Girl!" An average of one out of every five correct responses to probes received token reinforcement (i.e., a VR 5 schedule of reinforcement was used). Every correct probe trial was given social reinforcement.

Alternating Procedure

Training in the alternating procedure was similar to that of training one exemplar at a time, with the exception that two members (SM1 and SM8) belonging to one stimulus class was alternated with another two members (SM1 and SM8) belonging to a second stimulus class.

Training proceeded according to the following 8 step sequence:

1) Step 1 was the same as Step 1 in the OET procedure. Returns to this step from Steps 3 and 7 resulted in SM1 and SM8 being substituted for each other in the sequence.

2) Step 2 was the same as Step 2 in the OET procedure. Return to this step from Steps 4 and 8 resulted in SM1 and SM8 being substituted for each other in the sequence.

3) Step 3 was the same as Step 3 in the OET procedure except that SM8 was used instead of SM1.

4) Step 4 was the same as Step 4 in the OET procedure except that SM8 was used instead of SM1.

5) Step 5 was the same as Step 5 in the OET procedure.

6) Step 6 was the same as Step 6 in the OET procedure.

7) Step 7 was the same as Step 7 in the OET procedure except that SM8 was used instead of SM1.

8) Step 8 was the same as Step 3 in the OET procedure except that SM8 was used instead of SM1.

The above training procedure sequence was presented six times to equate the Alternating condition with the OET condition in terms of the number of training steps.

After successful completion of the training cycle a sixth time, the responses were said to have "reached criterion" and were considered known. The four remaining stimulus classes were then trained, two at a time, using the same procedure. During training all correct prompts were given social reinforcement in the form of "Good Girl!". Correct responses to probes received token reinforcement on a VR 5 schedule of reinforcement.

Post-Training Test for Learning

Following the completion of training and prior to the Generalization Test, a Post-Training Test was conducted. This procedure was similar to that used by Welch and Pear (1980) and Salmon et al. (1986) and consisted of a series of probe trials. With those stimulus classes trained in the OET condition, once the SM1 of each stimulus class had reached criterion, it was probed ten times. The order of presentation of each stimulus class was randomized and correct responses were reinforced on a VR 5 schedule. If a subject failed to respond correctly on a probe trial, a prompt trial was presented for that stimulus class. Regardless of whether the subject responded correctly or

incorrectly to the prompt, the experimenter proceeded to the next probe trial. Once this test was completed the experimenter proceeded to the Generalization Test.

Once the Generalization Test was completed, the six stimulus classes were retrained using the second member (SM8) of each class. The Post-test for learning following the completion of the above training was similar in that each stimulus class was randomly probed ten times, with each stimulus member being probed five times. For those stimulus classes which were trained using the alternating procedure the two members of each stimulus class were randomly probed five times. The results of this Post-Training Test gave the experimenter a percentage known for each trained stimulus member.

The Post-Training Test was important since prior to the Generalization Test it was necessary to establish that the stimulus classes were well known. If the percentage of correct responses on the Post-Training Test was high and the subject failed to generalize, one can say with greater confidence that the poor responding of the subjects during the Generalization Test does indeed reflect a lack of generalization. If the stimulus classes are not well known, one would not expect much generalization to occur. For the purposes of this research, 80% known or 8 out of 10 probes

answered correctly was considered adequate. In addition, on the Post-Training Test following the alternating condition and following the training of the second member during the OET condition each subject was allowed no more than one error to one of the trained stimulus class members. If a subject scored lower than 80% with any one stimulus class or made more than one error to any one stimulus member, that stimulus class was retrained and a second Post-Training Test was conducted prior to testing for generalization.

Generalization Test

The test for generalization was conducted in the training room after the Post-Training Test for learning was completed. The procedure used for testing for generalization was identical to that employed during the Laboratory Test. The modifications of the Laboratory Test that were made for Dawn and Esther were also made for the Generalization Test with these children.

Dependent Variables

The dependent variables which were examined in the research were:

- 1) Average number of minutes to reach criterion for each training stimulus.
- 2) Percent of correct responses for all testing stimuli during the Generalization Tests.

- 3) Error Analysis: An analysis of the errors made during the Generalization Test by each subject was undertaken. A subject could have made three types of errors: a) he or she could overgeneralize (use one of the training names incorrectly to label one of the trained stimulus classes), b) use a name not trained in the research to label a trained stimulus class, and c) fail to emit any response or reply that they did not know the name of the object.

Interobserver Reliability

All training sessions were recorded on audio-tape and approximately 1/6 of these were played back to an independent observer who was familiarized with the experimenter's criteria for what was considered a correct or incorrect response. In addition, all of the Post-Training and Generalization Tests were recorded and approximately 1/3 were randomly selected and scored by an independent observer.

Reliability coefficients were obtained by calculating the ratio of agreements to agreements plus disagreements on those responses that the experimenter scored as correct and the ratio of agreements to agreements plus disagreements on those responses that the experimenter scored as incorrect.

The reliability coefficients are shown in Table 1. The reliability coefficient ranged from .85 to 1.0, with many showing perfect agreement.

Procedural Reliability

All training sessions were video taped and approximately 1/6 of the session tapes were viewed by an independent observer and scored as to how well the experimenter followed the training protocol according to the following criteria: 1) whether the appropriate stimulus class was presented, 2) whether the appropriate member of a stimulus class was presented, 3) whether the training sequence was adhered to, i.e., the correct presentation of prompts and probes, and 4) whether token and social reinforcement were delivered according to the appropriate schedule of reinforcement. The reliability checks, during training, indicated that: 1) the wrong stimulus class was presented on three occasions; 2) the wrong member of a stimulus class was presented on four occasions; 3) the training sequence of prompts and probes was incorrectly presented on one occasion; and 4) the presentation of token reinforcement failed to occur on two occasions.

In addition, all testing and generalization sessions were video taped and approximately 1/3 of these were viewed by an independent observer and scored as to how well the

Table 1

Interobserver Reliability Coefficients for Experiment 1

Child	Procedure					
	Training		Post-Training		Generalization	
	C*	I**	C	I	C	I
Dawn	1.00	.98	1.00	.85	.98	.98
Esther	.97	.96	1.00	1.00	.99	1.00
Kathy	.95	.97	1.00	.90	1.00	1.00

* C = Reliability coefficient for those responses the experimenter scored as correct.

** I = Reliability coefficient for those responses the experimenter scored as incorrect.

experimenter followed the training protocol according to the following criteria: 1) whether the appropriate stimulus class was presented, 2) whether the appropriate member of a stimulus class was presented, 3) whether the training sequence was adhered too, i.e., the correct presentation of probe and instruction following trials, and 4) whether token and social reinforcement were delivered according to the appropriate schedule of reinforcement.

These reliability checks indicated that: 1) the appropriate stimulus class was always presented; 2) the appropriate member of a stimulus class was always presented; 3) the training sequence was adhered to perfectly, i.e., the presentation of instruction following trials, known probe and unknown probe trials; and 4) on one occasion an unknown probe trial was socially reinforced during the generalization test and token reinforcement failed to occur on three occasions.

Results

The data for Dawn and Esther are presented across the four phases of the research. In phase 1, where one exemplar at a time was trained, the Laboratory and Generalization Tests were conducted by interspersing the trained stimulus classes with instructions. As will be made clear, Dawn responded poorly during the Generalization Test. It was

felt that she may have discriminated between probe trials which were never reinforced and instruction following trials which were occasionally reinforced. As a result the procedure for the Laboratory and Generalization Tests were changed for both Dawn and Esther and in phases 2, 3, and 4, unknown stimulus classes were interspersed with known stimulus classes. This change lessened the probability that Dawn or Esther would discriminate between reinforced and nonreinforced trials.

Figure 2 shows the mean number of minutes to reach criterion for each stimulus class for the three children. These calculations also include retraining time for those stimulus classes that had to be retrained as a result of failing to meet the criterion specified by the Post-Training Test. For the first condition where the exemplars were trained one at a time, the amount of time required to train each of the two stimulus members was added together. In general, both Dawn and Esther reached criterion in a relatively short period of time. Criterion was reached slightly faster when two exemplars were trained together. For Kathy, training was time consuming. Because of this it was possible to complete only two phases with her. While training two exemplars together was somewhat faster, it is impossible to say how reliable this effect is given that she only completed two phases of the research.

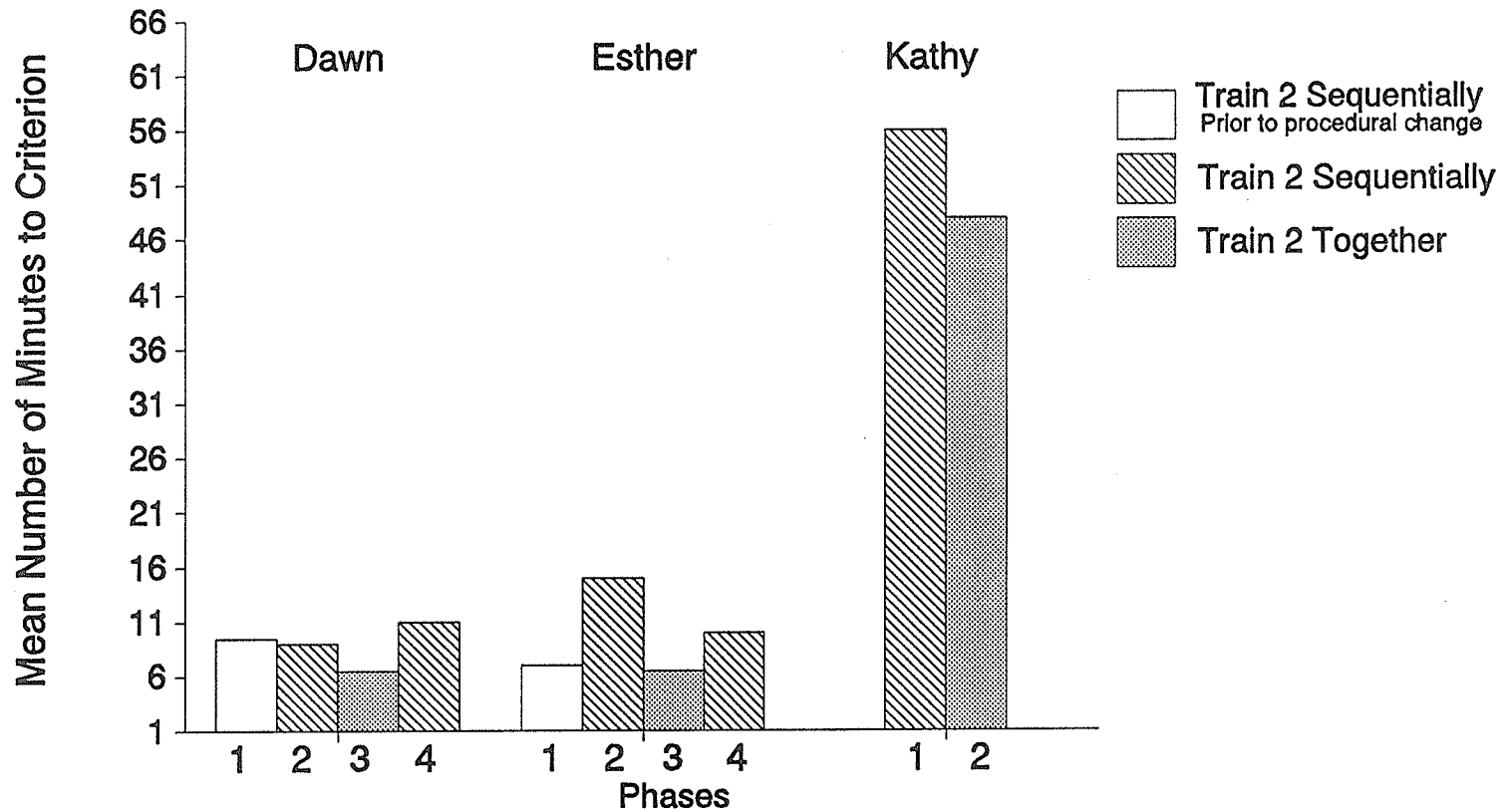


Figure 2 The mean number of minutes to reach criterion for each stimulus class after training two exemplars, one at a time (sequentially) or after training two exemplars together in Experiment I.

Figure 3 shows the amount of generalization for Dawn across the four phases of the experiment. The black dots indicate the results of the Post-Training Tests. Over the course of the research the names of five stimulus classes had to be retrained with Dawn. In the first phase after training one exemplar, generalization to untrained exemplars was low at approximately 10%. After training with the second exemplar, generalization to the untrained exemplars was found to be even lower at approximately 3%. Dawn often failed to respond to the testing stimuli during the Generalization Test.

Subsequently a second OET phase was conducted with training after one exemplar showing approximately 15% generalization to the testing stimuli. However, after training with the second exemplar there was a substantial increase in the amount of generalization to approximately 54%. In phase 3, after training both exemplars together, the percentage of generalization showed a small decline to approximately 46%. The following OET phase showed generalization with Dawn to be at approximately 23% after training with one exemplar. Following the training of a second exemplar, generalization increased to approximately 40%.

Dawn

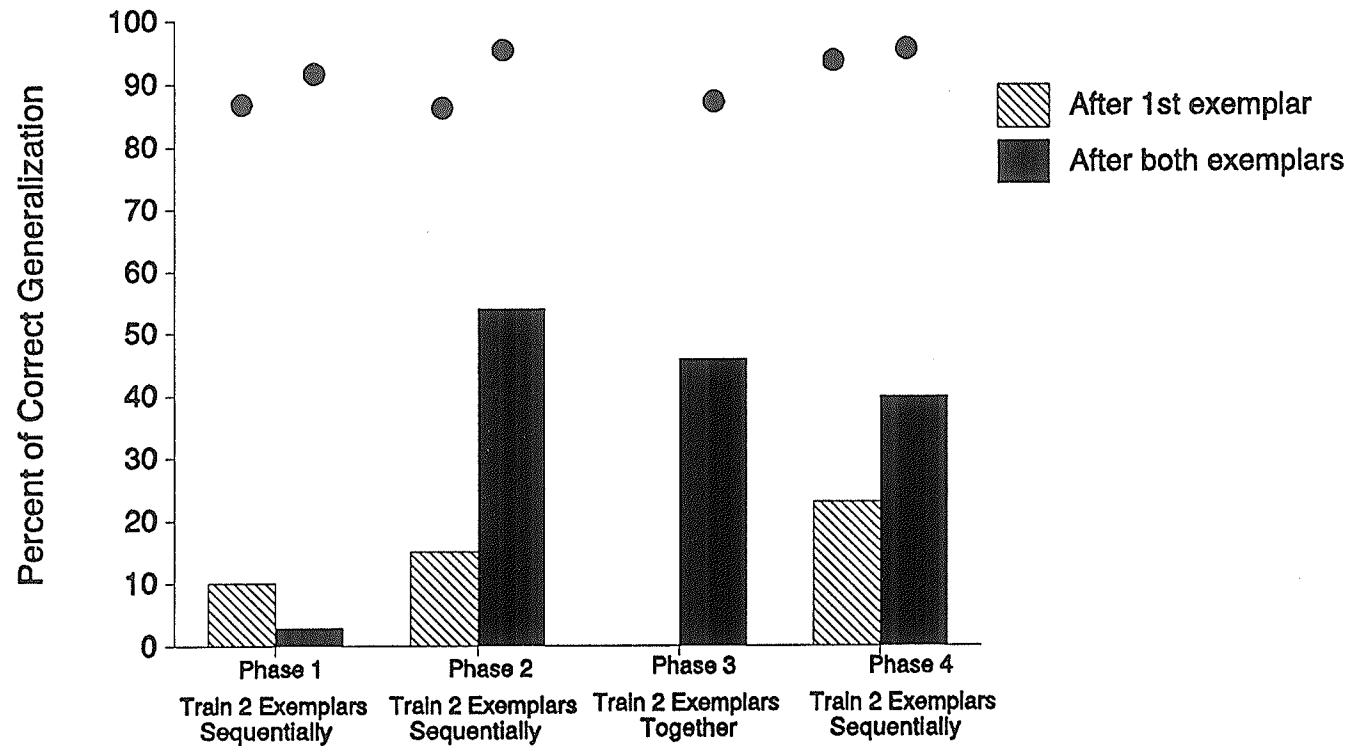


Figure 3 The percentage of correct generalization to testing stimuli after training two exemplars one at a time (sequentially), or after training two exemplars together. The data from each phase is presented as percent correct generalization for testing stimuli after one exemplar is trained, or after both exemplars are trained. The black dots indicate the percentage known on the post-training test.

Figure 4 shows the amount of generalization with Esther across the four training phases. Five stimulus classes had to be retrained with Esther. In the initial phase, after training with one exemplar, generalization to the testing stimuli was at approximately 18%. Training with the second exemplar resulted in a large increase in generalization to approximately 79%. However, because of the initial problem with Dawn and in order to be consistent in terms of the testing procedure, the Laboratory and Generalization Test with Esther was changed in the same manner as with Dawn and the phase was replicated. After training with only one exemplar generalization to the testing stimuli was fairly high at approximately 43%. Training with the second exemplar produced a small increase in generalization to approximately 56%.

Training in the third phase using the alternating procedure produced almost the same amount of generalization as that found in the second phase, approximately 54%. On returning to the OET phase, training with the first exemplar produced only about 8% generalization to the testing stimuli. Training with the second exemplar produced a moderate increase in generalization to approximately 32%.

Esther

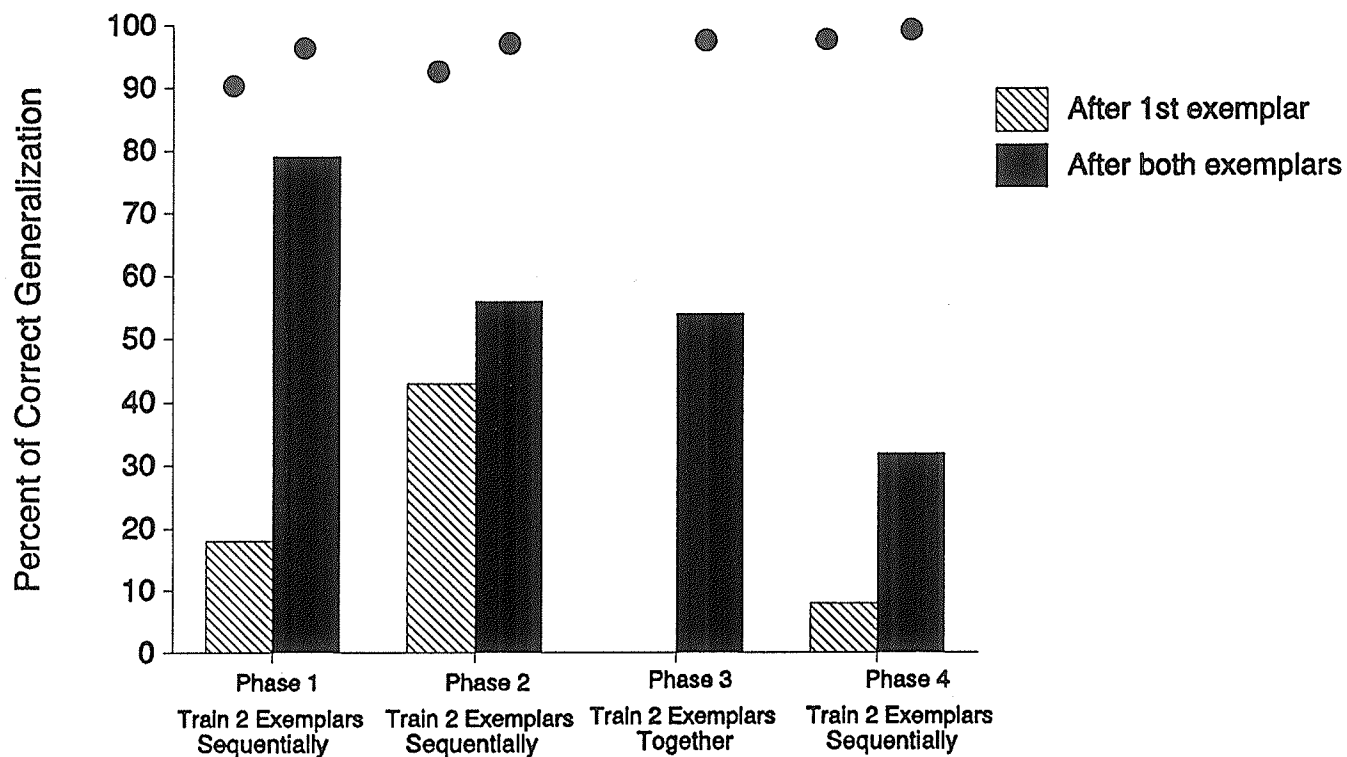


Figure 4 The percentage of correct generalization to testing stimuli after training two exemplars one at a time (sequentially), or after training two exemplars together. The data from each phase is presented as percent correct generalization for testing stimuli after one exemplar is trained, or after both exemplars are trained. The black dots indicate the percentage known on the post-training test.

Figure 5 shows the amount of generalization found with Kathy. She was extremely slow to learn and because of this the school year ended before a second OET phase could be completed. Although initially the procedural changes made with Dawn and Esther were to be made with Kathy, because of her slow learning rate these changes were not made. Thus for Kathy the Laboratory and Generalization Tests remained unchanged and testing was conducted using probes interspersed with instructions.

One stimulus class had to be retrained with Kathy. In the first phase after training with one exemplar she showed approximately 46% generalization to the testing stimuli. Following the training of the second exemplar generalization increased to approximately 70%. In the second phase after training with the alternating procedure, the amount of generalization found with the testing stimuli was similar at approximately 66%.

Table 2 shows the results of the error analysis for the three children. Dawn and Esther showed very little overgeneralization to other training and testing stimuli. As well, they seldom used any other name when presented with an object which they could correctly identify. In the majority of instances they either failed to respond or what

Kathy

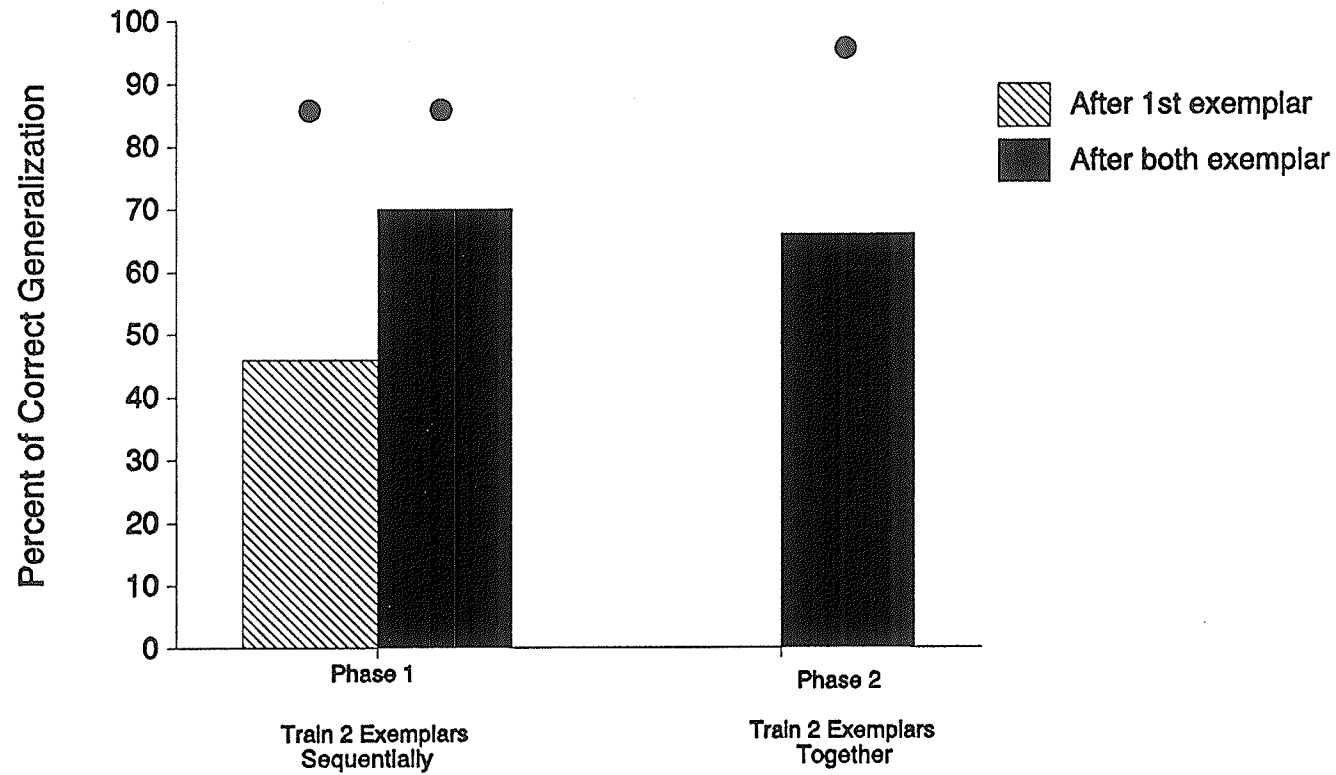


Figure 5 The percentage of correct generalization to testing stimuli after training two exemplars one at a time (sequentially), or after training two exemplars together. The data from each phase is presented as percent correct generalization for testing stimuli after one exemplar is trained, or after both exemplars are trained. The black dots indicate the percentage known on the post-training test.

was more probable said that they "didn't know". Katherine, on the other hand, showed a great deal of overgeneralization. She would often use the same name to incorrectly name an object. She also initially used a name other than that trained to label an object. However, this tended to decline following the training of the second exemplar in the first phase and in the second phase. She rarely failed to emit a response.

Table 3 shows the number of good medium and poor examples used with the three children (cf. Mervis & Pani, 1980). A total of 35 stimulus classes were trained. Nine of these were deemed by the raters to contain all good examples of a category. Twelve either contained all good or moderate examples of a category. Five contained good and poor examples of a category. The other nine contained a combination of good, moderate and poor examples of a category. What is perhaps most interesting and, as the table makes clear, most striking is the fact that 82% of all the objects used as either training or testing stimuli were rated as good examples of a particular category. Ten percent were rated as medium examples and eight percent as poor examples of a category.

TABLE 3

Number of Good, Medium, and Poor Examples of a Category for Experiment 1

<u>Type of Examples</u>	<u>Number of Examples</u>	<u>Percentage of Examples</u>
Good	231	82%
Medium	28	10%
Poor	21	8%

Table 4 shows the amount of generalization found with each child regardless of the training condition after training with two good examples, one good and one medium example, or one good and one poor example of a category. The data was combined from the two conditions since neither condition proved to be superior in terms of promoting generalization. The numbers in brackets are the number of categories (stimulus classes) trained using each of these three conditions. Given Dawn's low rate of responding in the first phase, it was excluded from these calculations. The first training exemplar, chosen by the experimenter, was always rated as a good example of a category. Of the 24 categories which contained good, medium, and/or poor examples of a category, 16 had an object rated as a medium or poor example of a category as the second training exemplar.

Table 4 clearly reveals that when training two good examples, the overall amount of generalization to the testing stimuli was not higher than that found when training included either one good or one medium example or one good or one poor example of a category. Indeed the most generalization found with all three children was when one good and one poor example was trained.

TABLE 4

Generalization Results from the Best Example Theory of
Categorization for Experiment 1

Child	Type of Examples Trained		
	G + G	G + M	G + P
DAWN	49% (8)	24% (5)	62% (5)
ESTHER	48% (11)	49% (8)	72% (5)
KATHY	58% (6)	67% (3)	85% (3)

G + G = Two Good Examples Trained

G + M = One Good Example and One Moderate Example Trained

G + P = One Good Example and One Poor Example Trained

() = The number of stimulus classes trained with each type
of examples

Discussion

The results from the first experiment indicate that both training procedures were equally effective in teaching the names of stimulus classes and promoting generalization. Both Dawn and Esther reached criterion in a relatively short period of time. The alternating procedure would seem to be a slightly faster means of teaching the names of the stimulus classes. This difference, while not large, may be significant when one is training a large pool of items. While Kathy took much longer to reach criterion than the other two children, the alternating procedure also proved to be less time consuming in terms of training the names of the stimulus classes. In the long run, if one were training a large number of stimulus classes the differences may be quite large. However, as the first phase was not replicated with Kathy, no unequivocal conclusions can be drawn.

Both training procedures proved to be equally effective in terms of promoting generalization. While the results with the alternating procedure replicate the finding of Salmon et al. (1986), it was not found to be superior to the OET condition. Once the procedural changes were made with Dawn, generalization increased substantially. Thus, she may well have initially discriminated between reinforced and nonreinforced trials. With the exception of the first phase

with Dawn, there was for both Dawn and Esther an overall decline in correct responding over the four phases of the research. The reason for this is unclear; it may reflect some weakening of the token reinforcement contingency. However, the children continued to acquire the names of the stimulus classes at a fairly consistent rate over the course of the research. In addition, although some stimulus classes had to be retrained, the results of the Post-Training Test were always high. Moreover the children were never reluctant to come to the training and testing sessions.

In conclusion both training procedures were equally effective in promoting generalization. However, from a cost-effective view, the training of two exemplars near the same point in time may represent a more efficient means of training stimulus classes, assuming that at least two exemplars have to be trained before sufficient generalization is noted.

The error analysis clearly reveals that Dawn and Esther made few naming responses (overgeneralizations or otherwise) to the testing stimuli. Indeed most of their responses took the form of "don't know". However, with Debbie there was a shift in her behavior from the first OET phase where she often refused to respond to the last three phases where she

was much more likely to respond "don't know" to a stimulus member she could not correctly name. This finding would seem to indicate that Dawn was discriminating between reinforced and non-reinforced trials during the first phase and that the procedural changes made were effective in alleviating this concern. The results of the error analysis with Kathy were quite different, in that she was much more likely to overgeneralize a response. Often she would use the same name each time she was presented with a stimulus member she could not name. She was the only child in the research diagnosed as being autistic, and, while she was fairly vocal, much of her behavior was under poor stimulus control.

The results from the "best example" ratings do not seem to support the best-example theory, as the training of one good and one poor example of a category produced as much or more generalization than did the training of two good examples. Caution is certainly necessary in interpreting these results given that most of the stimulus members (82%) were rated as good examples of a category. It may be that the training of one good example of a category may produce more generalization than the training of one poor example. Given that most of the stimuli were chosen on a rather basis with no attempt to determine whether they were good, medium, or poor examples of a category prior to conducting the

research indicates that most members of a category or a stimulus class may well be classified as good examples of a category and the number of medium and poor examples may, in general, be rarely found in a child's natural environment.

Since the alternating procedure was found to be somewhat superior to the OET procedure in terms of the amount of time it requires to reach criterion, it was chosen for Experiment 2. Also, with this procedure the Generalization Test occurs on only one occasion within each phase. Experiment 2 examined the issue of diversity of exemplars. When attempting to promote generalization the question is whether one should train exemplars that are similar to one another or whether one should train exemplars that reflect greater diversity. In this study, in one condition two members of a stimulus class that were similar to one another were trained and generalization to similar and different stimulus members of the same stimulus class was assessed. In a second condition, two stimulus members that were different from one another were trained and again generalization was assessed to similar and different stimulus members of the same stimulus class.

Experiment 2

Method

Subjects

Three subjects took part in Experiment 2. Two of the subjects (Lillian and Cindy) were enrolled in the day care program at a training and treatment facility, the St. Amant Centre in Winnipeg, Manitoba. The third (Dan) participated in the research at his home in Edmonton, Alberta. Prior to commencing the research, letters were sent to the parents of the children seeking permission to employ their child in the research.

Lillian was 4 years old at the time she participated in the present research. She had been diagnosed as being developmentally delayed. At age 3-10 her receptive and expressive language abilities were assessed at 3-0 on the Sequential Inventory of Communication Development - Revised. On the Peabody Picture Vocabulary test she scored at age 2-10.

Cindy was 3-6 at the time she participated in the research. Dan was 2-8 at the time he participated in the research. Both Cindy and Dan were developmentally normal and no formal assessment of their language abilities was conducted.

Setting and Apparatus

Training sessions for Lillian and Cindy were conducted in a room in the research area of the psychology department at the St. Amant Center. The child and experimenter sat opposite one another at a table. On the table a console timed the length of a trial and the inter-trial interval. The trial lasted for 8 seconds or until the child emitted a response, whichever came first. The inter-trial interval was 5 seconds. On a shelf beside the table a clock timed the length of each session. All sessions were 15 minutes long and each was videotaped. The training and testing stimulus were placed on the floor within easy reach of the experimenter but out of the children's sight.

Sessions with Dan were conducted in a room in his home in Edmonton, Alberta. Dan and the experimenter again sat opposite one another. The length of the trial and inter-trial intervals was the same as used with Lillian and Cindy, as was the length of each session. However, a video recorder was not available and all sessions were audiotaped.

Training and Testing Stimuli

As an in Experiment 1, 40 stimulus classes were selected prior to the start of the research. Many of the stimulus class members were the same as those used as in the first experiment. Within each stimulus class eight stimulus

members were selected for training and testing purposes. These were selected on the basis of three separate ratings obtained with introductory psychology students. Prior to the ratings, the experimenter acquired anywhere from eight to fourteen members of a stimulus class. This was done with all forty stimulus classes. Many of the stimulus members were from the first experiment, while others were purchased specifically for this experiment. The number of stimulus members initially used in the ratings was determined largely by their cost.

The rating for the second experiment differed considerably from the first experiment. In Experiment 1, the ratings were designed to rank order the different stimuli of each stimulus class to determine which two stimuli were least similar to one another. In Experiment 2, the ratings were designed to identify two different groups of stimuli within a stimulus class, with the stimuli in each group being similar to one another.

For the first rating ten introductory psychology students were asked to place each stimulus member of a stimulus class into one of two groups (1 or 2) using similarity as a basis for grouping them. The rating instructions required that the stimuli placed in the first group be similar to one another, yet different from the

stimuli placed in the second group which also had to be similar to one another. If in the opinion of the rater a stimulus did not belong with either group the raters were instructed to place it in a miscellaneous group. The raters were informed that there was no set number of members that had to be placed in either group. (See Appendix E for the rating instructions.)

Following this initial rating of each stimulus class, some of the members placed in one group were transposed in order to balance out the ratings. For example, if one rater for the stimulus class "BADGE" had placed the stimuli labelled A, B, C, D, into group 1, and the stimuli labelled E, F, G, H into group 2 and a second rater had placed the stimuli labelled A, B, C, D, into group 2 and the stimuli labelled E, F, G, H into group 1 the second rating was simply transposed with the stimuli labelled A, B, C, D, E, placed in group 1 and the stimuli labelled E, F, G, H in group 2.

If more than three raters disagreed on the assignment of a stimulus to a particular group it was discarded. Following this first rating, for some stimulus classes, fewer than four stimulus members were left in a group. Subsequently, new members were added to these stimulus classes and all forty were rated a second time using a

second group of ten introductory psychology students. The rating instructions given to this group of students was identical to those given to the first group of raters. Again, if more than three of the raters could not agree on which group a member of a stimulus class should be placed in, it was discarded.

This second rating produced enough agreement such that for each stimulus class four or more members remained in each of the two groups. If there were more than four members in a group, the experimenter randomly chose the four that would be retained for the experiment. Once eight stimulus members, four in group 1 and four in group 2, had been selected for all forty stimulus classes a final rating was conducted. The purpose of this rating was to further confirm the experimenter's selection of stimulus members. It differed from the first two ratings in that a third group of ten introductory psychology students were given a forced choice. For all forty stimulus classes, the raters were asked to place the eight members of a class, chosen from the second rating, into two groups. The ratings were made again on the basis of similarity. However, the raters were forced to place four stimulus members in group 1 and four stimulus members in group 2. It was hypothesized that there should be high agreement between this forced choice rating and the second rating. (See Appendix F for the forced choice rating

instructions.) The forced choice rating produced 100% agreement for 37 of the 40 stimulus classes, 90% for two stimulus classes and 70% for one stimulus class.

Prior to commencing the research, the two groups of four stimulus members that made up a stimulus class were randomly assigned to either group 1 or 2. This procedure was conducted with all forty stimulus classes. Following this, two members in group 1 and one in group 2 were randomly chosen as a training stimuli for each stimulus class.

In addition, the new stimuli that were acquired for this research were rated as to whether they were good, moderate, or poor examples of a stimulus class (cf. Mervis & Pani, 1980). (The results of these ratings can be found in Appendix G.) On a post hoc basis, raters were asked to rate each of the three training stimuli for each stimulus class in relation to the other members in a stimulus class, using similarity or shape as a basis for rating them. The ratings were conducted for all 40 stimulus classes. A total of six groups of ten raters each were used for these ratings. Each group rated one of the training stimuli in relation to the other seven stimulus members using either similarity or shape as a basis for rating them. The purpose of these

ratings was to further confirm the experimenter's assignment of stimulus members to either group 1 or 2.

The placement of the stimuli into groups provides only nominal information. The post hoc ratings provide ordinal information about the stimulus members of each stimulus class. The ratings should reveal that the stimulus members belonging to one group should be rated as being more similar to the training stimulus (stimuli) belonging to the same group than those stimulus members that belong to the other group. The shape ratings were conducted to see if there was any correspondence between the amount of generalization noted and how similar the members of a stimulus class were in terms of shape. (See Appendixes H and I for the similarity and shape rating instructions. The data for these ratings can be found in Appendix J.)

Experimental Design

For all three children the stimulus classes were trained using a replication design. Initially two replications were planned. However, after the preliminary assessment it was discovered that neither Lillian or Cindy had the required number of unknown stimulus classes and as a result only one replication was conducted. In each phase, each child was trained to name six stimulus classes. In phase 1, two members from each stimulus class were trained.

These were the ones that had been rated as being similar to one another. These two stimulus members were always chosen from group 1. In phase 2, two members from each stimulus class were trained using two stimulus class members that had been rated as being different from one another, one from group 1 and one from group 2. Phase 3 was a replication of phase 1. For Dan only, phase 4 was a replication of phase 2.

Preliminary Procedures

Each subject proceeded through the same preliminary steps as in Experiment 1. As in Experiment 1 each subject was worked with on an individual basis.

Overview of Training and Testing Procedures

Each subject proceeded through the same sequence of testing and training as was used in Experiment 1. The alternating procedure was used throughout Experiment 2. In addition, for the Generalization Test, instructions were interspersed with probes in the same manner as had been originally proposed in Experiment 1. Primary and social reinforcement was also administered in the same manner. During training and the Post-Training Test all correct responses to probes and prompts were given social reinforcement. Correct responses to probes were also reinforced with primary reinforcement using a VR 5 schedule

of reinforcement. Primary reinforcers consisted of a choice between pieces of "Smarties", chips, "Cheezies", or 1/2 teaspoon of ice cream.

Dependent Variables

The dependent variables in Experiment 2 were identical to those of Experiment 1.

Interobserver Reliability

Table 5 shows the results of the interobserver reliability checks. IORs were conducted in the same manner to that employed in Experiment 1. Coefficients ranged from .9 to 1.0, with many instances of perfect agreement.

Procedural Reliability

Procedural reliability checks were conducted in the same manner as in Experiment I. The reliability checks during training, indicated that: 1) the wrong stimulus class was presented on two occasions; 2) the wrong member of a stimulus class was presented on two occasions; 3) the training sequences of prompts and probes was incorrectly presented on two occasions, and 4) the presentation of primary reinforcement failed to occur on one occasion.

The reliability checks, during the Post training and Generalization tests indicated that: 1) the appropriate

Table 5

Interobserver Reliability Coefficients for Experiment 2

Child	Procedure					
	Training		Post-Training		Generalization	
	C*	I**	C	I	C	I
Lillian	.98	.95	.99	.90	1.0	1.0
Cindy	1.0	1.0	1.0	.92	1.0	1.0
Dan	1.0	.90	.98	1.0	1.0	1.0

* C = Reliability coefficient for those responses the experimenter scored as correct.

** I = Reliability coefficient for those responses the experimenter scored as incorrect.

stimulus class was always presented; 2) the appropriate member of a stimulus class was always presented; 3) the training sequence was adhered to perfectly, i.e., the correct presentation of instruction following trials and probe trials, and 4) on two occasions a probe trial was socially reinforced during the generalization test and primary reinforcement failed to occur on two occasions.

Results

Figure 6 shows the mean amount of time it took all three children to reach criterion for each stimulus class. Lillian, in phase 1, averaged approximately 11 minutes for each stimulus class. In phase 2 this increased substantially to 19 minutes per stimulus class. Upon returning to phase 1, there was a further small increase to 22 minutes per stimulus class. The reasons for these increases is difficult to determine but may represent some fatigue factor with regard to the training procedure. For Cindy the average amount of time to reach criterion for each stimulus class, while showing some small increase, did not seem to differ significantly over phases. For Dan, the amount of time to reach criterion was fairly consistent across the four phases.

Figure 7 shows the amount of generalization found with Lillian over the three phases of the experiment. Each phase

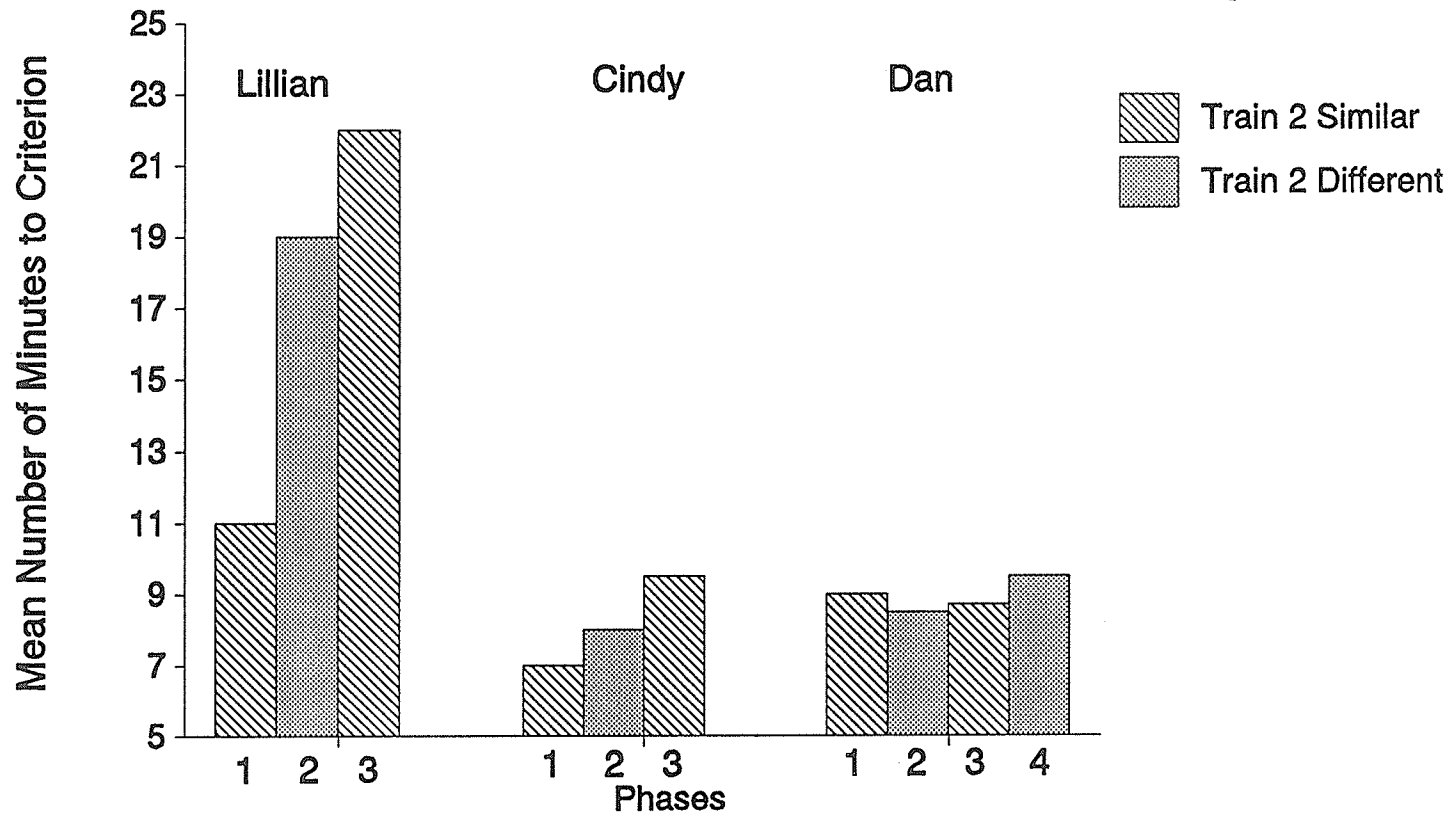


Figure 6 The mean number of minutes to reach criterion for each stimulus class after training two similar or two different exemplars in Experiment II.

Lillian

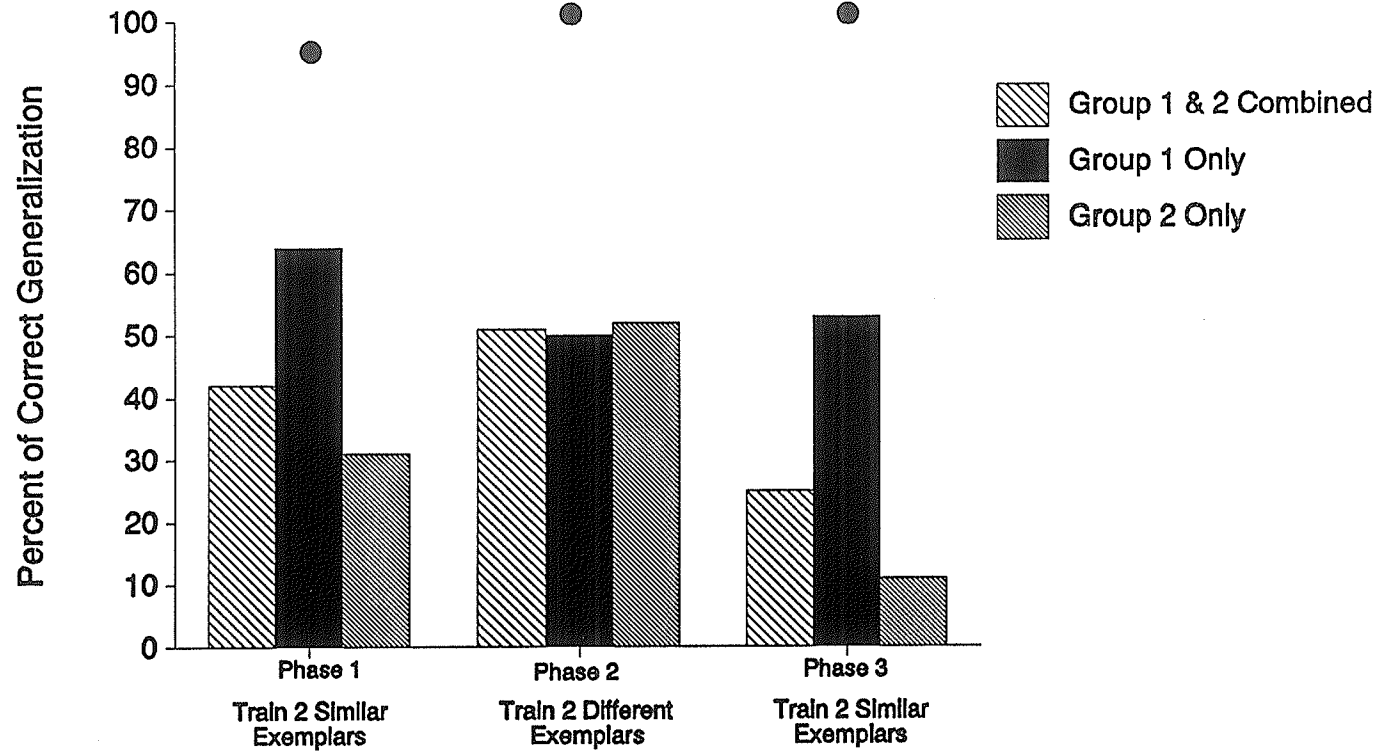


Figure 7 The percentage of correct generalization to testing stimuli following training with two similar exemplars from group 1, or after training with two different exemplars, one from group 1 and one from group 2. The data from each phase is presented as percent correct generalization for all testing stimuli (groups 1 and 2 combined), for the testing stimuli in group 1 alone, and for testing stimuli in group 2 alone. The black dots indicate the percentage known on the post-training test.

of the experiment is presented in terms of the amount of generalization found with all the testing stimuli (groups 1 and 2 combined), the amount of generalization found for those testing stimuli belonging to group 1, and the amount of generalization found with those testing stimuli belonging to group 2. The black dots indicate the results of Post-Training Test. Over the course of the research, two stimulus classes had to be retrained with Lillian.

In the first phase after training two similar stimulus members from group 1, generalization to all members from both group 1 and 2 was at approximately 41%. However, generalization to the testing stimuli belonging to group 1 was much higher at approximately 64%. With the testing stimuli belonging to the group 2, generalization was considerably lower at approximately 31%. In the second phase where one stimulus member was trained from group 1 and another from group 2, generalization to all testing stimuli was somewhat higher at 51%. Interestingly, generalization to the testing stimuli belonging to groups 1 and 2 was almost identical at 50% and 52% respectively. After replicating the first phase, generalization to all testing stimuli was considerably lower at approximately 26%. However, consistent with the previous phase 1 results, the amount of generalization found to the testing stimuli belonging to the group 1 was consistently higher at

approximately 55%. Generalization to the testing stimuli in group 2 was very low at only 11%.

Figure 8 shows the amount of generalization found with Cindy over the three phases of the research. Four stimulus classes had to be retrained with Cindy. In the first phase, after training two similar stimulus members from group 1, generalization to the testing stimuli belonging to groups 1 and 2 was at approximately 44%. Generalization to those testing stimuli belonging to group 1 was very high at approximately 83%. On the other hand generalization to the testing stimuli belonging to group 2 was considerably lower at approximately 25%. In the second phase, after training one stimulus member from group 1 and one from group 2, there was a large increase in the amount of generalization to all testing stimuli to approximately 87%. Consistent with the results from Lillian, generalization was evenly divided across the testing stimuli from groups 1 and 2 at 89% and 85% respectively. On replicating the first phase the amount of generalization to all testing stimuli fell considerably to approximately 45%. Consistent with the previous first phase and with the results found with Lillian, generalization to those stimuli belonging to group 1 was considerably higher at approximately 75% while generalization to group 2 was considerably lower at 31%.

Cindy

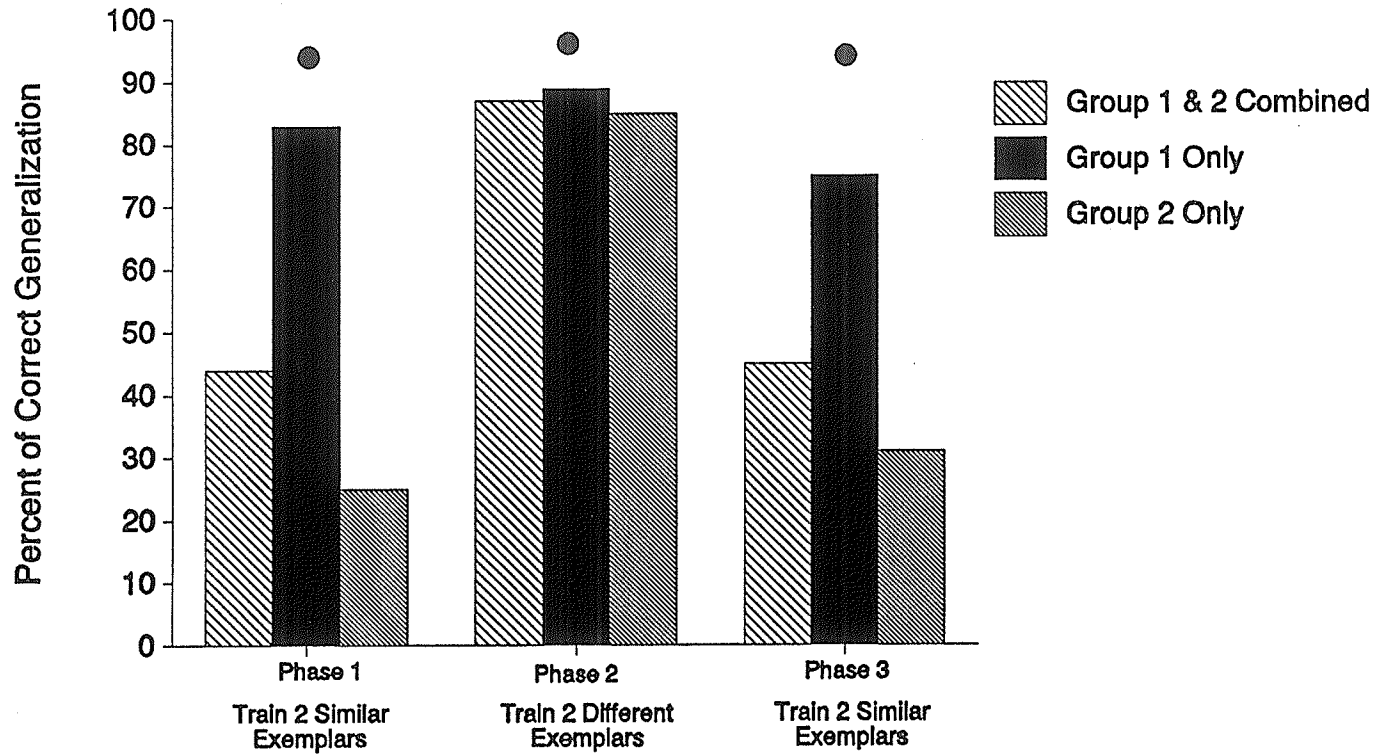


Figure 8 The percentage of correct generalization to testing stimuli following training with two similar exemplars from group 1, or after training with two different exemplars, one from group 1 and one from group 2. The data from each phase is presented as percent correct generalization for all testing stimuli (groups 1 and 2 combined), for the testing stimuli in group 1 alone, and for testing stimuli in group 2 alone. The black dots indicate the percentage known on the post-training test.

Figure 9 shows the amount of generalization found with Dan over the four phases of the research. Two stimulus classes had to be retrained with Dan. In the first phase, after training with two stimulus members from group 1, the amount of generalization to all testing stimuli was at approximately 43%. Generalization to the testing stimuli found in group 1 was at approximately 78%, while generalization to the testing stimuli belonging to group 2 was at approximately 26%. In the second phase, after training with one stimulus member from group 1 and one from group 2, the amount of generalization was high at approximately 87%. Generalization to the testing stimuli belonging to groups 1 and 2 was almost identical at 89% and 85% respectively. On replicating the first phase, the amount of generalization to all testing stimuli is higher than that found in the first phase at approximately 55%. Generalization to the testing stimuli in group 1 is again high at approximately 86%, while generalization to the testing stimuli in group 2 is much lower at approximately 38%. On replicating the second phase, generalization to all the testing stimuli was at 83%. Generalization to the members in groups 1 and 2 was also identical at 83%. The amount of generalization is masked by some difficulty Dan had with one stimulus class. In this phase one stimulus class had to be retrained. Although Dan was able to achieve

Dan

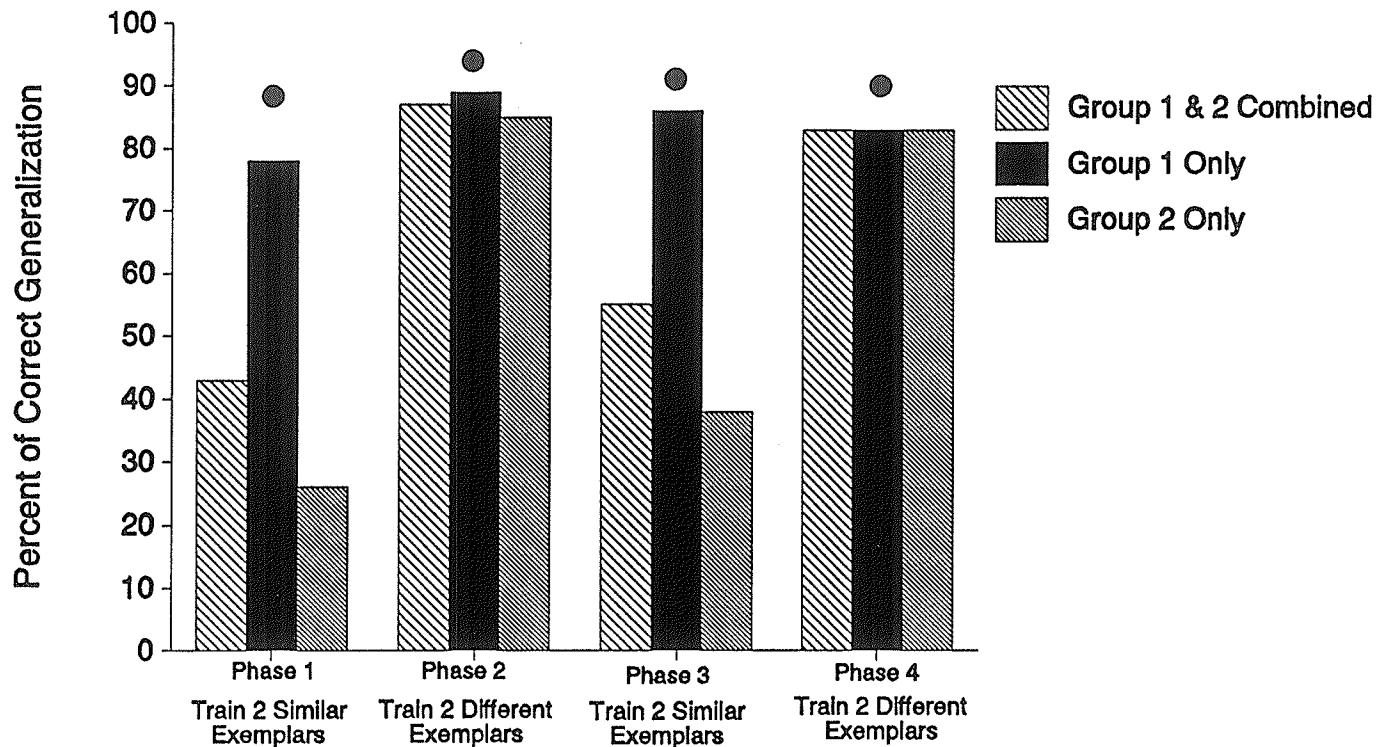


Figure 9 The percentage of correct generalization to testing stimuli following training with two similar exemplars from group 1, or after training with two different exemplars, one from group 1 and one from group 2. The data from each phase is presented as percent correct generalization for all testing stimuli (groups 1 and 2 combined), for the testing stimuli in group 1 alone, and for testing stimuli in group 2 alone. The black dots indicate the percentage known on the post-training test.

90% on the Post-Training Test, he did not respond to any of the training or testing stimuli on the Generalization Test. His performance with the other stimulus classes reflects perfect generalization.

Table 6 shows the results of the error analysis. The three children made very few overgeneralization errors. Both Lillian and Dan often used a name other than a training name to describe the testing members. Often this was the same name they used in the Laboratory Test to incorrectly label a stimulus member. All three children, when presented with a stimulus member they could not name typically responded with "don't know". They rarely failed to emit a response.

Table 7 shows the break down of the different stimulus classes in terms of good, medium, and poor examples of a category (cf. Mervis & Pani, 1980). A total of 26 stimulus classes were trained between the three children. Forty percent of the stimuli used in this research were added specifically for this experiment while the remainder were ones that had been used in the first experiment. As the table makes clear, despite the infusion of new stimuli into the experiment, the percentage of good, medium, and poor examples of a category were very similar to that found in the first experiment. Eighty-three percent, nine percent

TABLE 6

Error Analysis Data For Experiment 2

Phase *Type of Errors	Lillian			
	1	2	3	4
1	1	6	1	
2	52	42	37	
3	10	5	38	
	Cindy			
	1	2	3	4
1	2	1	1	
2	4	0	0	
3	54	13	58	
	Dan			
	1	2	3	4
1	6	5	0	5
2	28	6	26	0
3	28	3	23	13

*Type of Errors

- 1 Overgeneralization (use of one of the trained names incorrectly to label an object)
- 2 Use of a name not trained to label an object
- 3 Fails to emit a response or replies that she does not know the name

TABLE 7

Number of Good, Medium, and Poor Examples of a Category for Experiment 2

<u>Type of Examples</u>	<u>Number of Examples</u>	<u>Percentage of Examples</u>
Good	172	83%
Medium	19	9%
Poor	17	8%

and eight percent of the stimulus members were found to be good, medium and poor examples of a category respectively.

Table 8 shows the amount of generalization found after training in both conditions of the research. Only those stimulus classes trained using two good examples of a category are displayed in the table. If a good and medium or good and poor examples were trained these results were not considered in the calculations. The rationale for the above is that according to the best example theory the training of one or more good example should be an effective means of prompting generalization. The two training conditions studied in this research should be irrelevant, as long as both examples of a category have been rated as good examples of it. The table clearly reveals that the amount of generalization found depended on the training conditions and was not specifically affected by the training of two good examples. Lillian showed 33% generalization after training with two similar exemplars versus 64% after training with two different exemplars. Cindy showed 51% and 90% generalization respectively and Dan showed 52% and 85% generalization respectively.

Table 8

Percentage of Correct Generalization After Training Two Good Examples of a Category by Training Condition

Training Condition		
Child	Train Two Similar Stimulus Members	Train Two Different Stimulus Members
Lillian	33%	64%
Cindy	51%	90%
Dan	52%	85%

General Discussion

While the results of Experiment 1 indicate that generalization is not affected by training two exemplars separately or together, the results of Experiment 2 clearly indicate that diversity of exemplars promotes greater generalization than does similarity of exemplars. These results were consistent both within and between children. It is likely that the training of two diverse exemplars samples a broader range of relevant stimulus characteristics. On the other hand the training of two exemplars that are similar to one another may promote very tight stimulus control over a narrow range of stimulus characteristics. From a purely applied point of view the results would seem to qualify Stokes and Baer's (1977) suggestion of "training sufficient" exemplars. While "sufficient" exemplars is important, it is also necessary to consider the types of exemplars trained and whether they sample the broad range of stimulus characteristics needed to promote generalization. Consideration of this, prior to training, may result in a substantial reduction in the amount of training time necessary to promote optimal levels of generalization. The results fit nicely into Stokes and Osnes (1989) concept of "Train Diversely", but again qualify it in the same manner as above.

Lillian was the only handicapped child in Experiment 2 and it could be argued that these findings may not be reliable with other handicapped children. While the size of the effect with her was not as large as with the other children, the trend is clearly in the same direction. In addition Lillian also took longer to reach criterion. It may be that handicapped children are more in need of training with two or more similar exemplars within each group whereas normal children seem to acquire the concept after training with only one exemplar from each group.

What Experiment 2 does not demonstrate is whether the training of only two diverse exemplars would promote generalization to all members of a stimulus class. It is clearly possible that more than two diverse groups of similar stimuli can be identified within a stimulus class. The ratings certainly indicate that some stimulus members were not seen as being members of either group. The present results may add support to the quantal view of stimulus generalization (Bickel & Etzel, 1985). This view suggests that a stimulus-response relationship represents a unit that may or may not occur. Traditionally, stimulus generalization has been viewed as a continuous relationship between stimuli and responses. Thus, the second group of stimulus members for all intents and purposes may represent a new unit that has to be explicitly trained. Within any

one stimulus class there may be a number of different units.

In the future, research which addresses this issue may be of some value. If for a number of stimulus classes a third group can be reliably identified, it would be interesting to see if the training of one exemplar each from the first two groups would result in generalization to the third group. If it did not and the third had to be explicitly trained, this would add support to the quantal view of generalization. The results may also be interpreted in terms of summation of generalization gradients after training at two different points on the stimulus continuum (Kalish & Guttman, 1957).

It may perhaps be argued that the assignment of members to a group was flawed and that they did not represent distinct categories. However, the ratings obtained on a post hoc basis using both similarity and shape as relevant dimensions would seem to add further support to the assignment of members to their respective groups. What is also interesting about these ratings is that, for many of the stimulus members rated, the mean shape and similarity ratings are similar to one another. It may well be that when people are asked to use similarity as a basis for rating stimulus members, shape may be the dimension which is the primary focus of the rating.

The results of Experiment 2 would not seem to support the best example theory of categorization. When training with two good examples generalization should not, according to this theory, be affected by whether the stimulus members are similar or different from one another. However, the latter was clearly what influenced the children's responding. As with Experiment 1, most of the stimuli were rated as good examples of a category, despite the fact that 40% of the stimuli were specifically selected for this experiment. It is possible, given the wide exposure introductory psychology students have had to numerous objects, that what is considered a good example may be based more on the familiarity of the object than any other characteristic.

Caution is advised, however, in evaluating the implications of Experiment 2 for the best example theory, since the analysis was performed on a post hoc basis. Future research should examine this theory more closely in relation to the present research findings. For example, prior to commencing an experiment it would be desirable to group stimulus members on the basis of similarity and also rate them as to whether they are good, medium, or poor examples of a category. Ideally, one would want an equal number of good, medium, and poor examples of a category

within each group that has been designated on the basis of similarity.

What is perhaps most encouraging from this research is that adults seem to be able to reliably group members of the same stimulus class into at least two diverse groups. Once these groups have been identified, applied practitioners need only train one or two exemplars per group to maximize generalization across the largest number of members within a stimulus class.

References

- Anderson, S.R., Spradlin, J.E. (1980). The generalized effects of productive label training involving common objects. Journal of the Association for the Severely Handicapped, 5, 143-157.
- Bickel, W.K., Etzel, B.C. (1985). The quantal nature of controlling stimulus-response relations as measured in tests of stimulus generalization. Journal of the Experimental Analysis of Behavior, 44, 245-270.
- Bucher, B. (1973). Some variables effecting children's compliance with instructions. Journal of Experimental Child Psychology, 15, 10-21.
- Guttman, N., & Kalish, H.I. (1956). Discriminability and stimulus generalization. Journal of Experimental Psychology, 51, 79-88.
- Handleman, J.S., Powers, M.D., Harris, S.L. (1984). Teaching of labels: An analysis of concrete and pictorial representations. American Journal of Mental Deficiency, 88, 625-629.
- Hupp, S.C. (1980). Application of the best-example theory of categorization to concept development and communication training with severely handicapped students. Doctoral dissertation. University of Illinois at Urbana-Champaign.

- Hupp, S.C. Mervis, C.B. (1982). Acquisition of basic object categories by severely handicapped children. Child Development, 53, 760-767.
- Kalish, H.I., & Guttman, N. (1957). Stimulus generalization after equal training on two stimuli. Journal of Experimental Psychology, 53, 139-144.
- Keller, F.S., & Schoenfeld, W.N. (1950). Principles of Psychology. New York: Appleton-Century Crofts.
- Martin, J.A. (1974). The control of imitative and non-imitative behaviors in severely retarded children through "generalized -instruction following". Journal of Experimental Child Psychology, 11, 390-400.
- Martin, G.L., & Pear, J.J. (1988). Behavior Modification: What is it and how to do it. Third Edition. Englewood Cliffs, N.J.: Prentice Hall.
- Mervis, C.B., & Pani, J.R. (1980). Acquisition of basic object categories. Cognitive Psychology, 12, 496-522.
- Olenick, D.L., & Pear, J.J. (1980). Differential reinforcement of correct responses to prompts and probes in picture naming training with severely retarded children. Journal of Applied Behavior Analysis, 13, 77-90.
- Parsonson, B.S., & Baer, D.M. (1978). Training generalized improvisation of tools by preschool children. Journal of Applied Behavior Analysis, 11, 363-380.

- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), Cognitive Development and the Acquisition of Language. New York: Academic Press.
- Rowan, V.C., & Pear, J.J. (1985). A comparison of the effects of interspersal and concurrent training sequences on acquisition, retention and generalization. Applied Research in Mental Retardation, 6, 127-145.
- Salmon, D.J., Pear, J.J., & Kuhn, B.A. (1986). Generalization of object naming after training with picture cards and with objects. Journal of Applied Behavior Analysis, 19, 53-58.
- Spradlin, J.E. (1982). Stimulus classes and verbal behavior. Paper presented at the 8th Annual Convention of the Association for Behavior Analysis. Milwaukee, Wisconsin.
- Stokes, T.F., Baer, D.M. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.
- Stokes, T.F., Baer, D.M., Jackson, R.L. (1974). Programming the generalization of a greeting response in four retarded children. Journal of Applied Behavior Analysis, 7, 599-610.
- Stokes, T.F., & Osnes, P.G. (1989). An operant pursuit of generalization. Behavior Therapy, 20, 337-355.

Welch, S.J., Pear, J.J. (1980). Generalization of naming responses to objects in the natural environment as a function of training stimulus modality with retarded children. Journal of Applied Behavior Analysis, 13, 629-644.

Appendix A

Rating Instructions For Experiment 1

Each stimulus class was rated separately. Raters worked independently of one another. The following instructions were read to each rater:

On the table in front of you are 40 different categories of objects. Within each category there are 8 objects which differ from one another along a number of dimensions. However, we would typically call each object by the same name (e.g. kettle, clamp, masher). Each object within a category has been labelled with letters A-H. I would like you to first locate the object labelled with the letter "A". I would then like you to rank order the remaining 7 objects in relation to the "A" object using similarity as a basis for ranking them. The objects should be ranked from most to least similar to the A object.

On the data sheet provided you will note the name of each object, the letters A-H and a blank space below each letter. Below the letter A the number "1" has been placed. Once you have rank ordered the 7 objects the appropriate number should be placed in the space provided below the appropriate letter. For example if for a certain category you deem the "D" object to be most similar to the "A" object, the number 2 should be placed below it. If the "F" object is deemed least similar the number 8 should be written in the space below the letter F. The remaining 5 objects should be assigned numbers in a similar manner. If for some reason you believe an object does not belong to the category which it has been assigned place a "X" below the appropriate letter. Do you have any questions?

Appendix B

Final Ratings of Stimulus Classes For Experiment 1

	Stimulus Member	Raters									
		1	2	3	4	5	6	7	8	9	10
Badge	A	1	1	1	1	1	1	1	1	1	1
	B	3	3	3	3	3	3	3	3	3	3
	C	4	4	5	5	5	4	5	5	5	5
	D	7	8	7	7	7	8	7	7	7	7
	E	6	7	6	6	6	7	6	8	6	6
	F	2	2	2	2	2	2	2	2	2	2
	G	8	6	8	8	8	6	8	6	8	8
	H	5	5	4	4	4	5	4	4	4	4
Beater	A	1	1	1	1	1	1	1	1	1	1
	B	2	5	2	2	2	2	2	2	2	2
	C	7	7	8	7	8	8	8	8	8	7
	D	6	8	6	8	6	6	6	6	6	4
	E	5	4	5	4	5	5	5	4	5	6
	F	3	2	4	5	4	3	3	5	3	5
	G	4	3	3	6	3	4	4	3	4	3
	H	8	6	7	3	7	7	7	7	7	8
Bracelet	A	1	1	1	1	1	1	1	1	1	1
	B	7	8	8	8	8	8	7	8	8	8
	C	2	2	2	2	2	2	2	3	2	2
	D	4	4	4	3	4	4	4	4	4	4
	E	3	3	3	4	3	3	3	2	3	3
	F	5	6	7	6	6	7	6	6	7	6
	G	8	7	6	7	7	6	8	7	5	7
	H	6	5	5	5	5	5	5	5	6	5
Bowl	A	1	1	1	1	1	1	1	1	1	1
	B	3	3	2	3	2	2	3	2	2	3
	C	2	4	3	4	3	4	2	5	3	2
	D	4	2	5	2	4	5	4	3	4	4
	E	6	5	4	5	6	3	5	6	7	5
	F	5	6	6	6	5	6	6	4	8	6
	G	7	7	7	8	8	7	7	7	6	7
	H	8	8	8	7	7	8	8	8	8	8

Candle	A	1	1	1	1	1	1	1	1	1	1
	B	4	2	5	5	4	2	4	5	4	5
	C	2	5	2	3	3	3	2	2	2	2
	D	3	4	3	2	2	4	3	3	3	3
	E	5	6	7	8	7	7	7	7	7	7
	F	7	7	8	7	8	8	8	8	8	8
	G	8	3	6	4	5	6	5	4	5	6
	H	6	8	4	6	6	5	6	6	6	4
Chain	A	1	1	1	1	1	1	1	1	1	1
	B	4	3	4	3	3	4	4	3	3	3
	C	8	7	8	7	8	8	8	8	8	8
	D	6	6	7	8	7	7	7	7	6	6
	E	3	8	3	5	5	6	6	6	7	7
	F	2	2	2	2	2	2	2	2	2	2
	G	7	5	6	6	6	3	3	4	4	5
	H	5	4	5	4	4	5	5	5	5	3
Clamp	A	1	1	1	1	1	1	1	1	1	1
	B	3	8	8	8	8	8	8	8	8	8
	C	7	7	4	4	7	7	7	7	6	7
	D	5	5	6	6	5	5	5	5	5	4
	E	2	2	2	2	3	2	2	2	2	2
	F	6	6	5	5	6	6	6	6	7	5
	G	4	4	3	3	4	4	4	4	4	6
	H	8	3	7	7	2	3	3	3	3	3
Clip	A	1	1	1	1	1	1	1	1	1	1
	B	2	2	3	2	2	2	3	3	2	2
	C	3	3	2	3	4	3	2	4	3	4
	D	5	4	5	5	3	4	4	2	4	3
	E	4	6	4	4	6	5	5	5	6	5
	F	6	5	7	6	5	7	6	6	5	7
	G	7	7	6	7	7	6	7	7	7	6
	H	8	8	8	8	8	8	8	8	8	8
Clothespin	A	1	1	1	1	1	1	1	1	1	1
	B	4	6	3	4	3	3	3	4	6	3
	C	8	7	8	8	8	8	7	7	4	8
	D	2	2	2	2	2	2	2	2	2	2
	E	6	5	7	5	5	5	5	6	8	5
	F	7	8	4	7	7	7	8	8	5	7
	G	5	4	6	6	4	6	6	5	7	4
	H	3	3	5	3	6	4	4	3	3	6

Coffeepot	A	1	1	1	1	1	1	1	1	1	1
	B	4	3	4	3	5	5	5	4	3	5
	C	8	6	8	7	8	8	8	7	8	8
	D	6	8	6	8	6	7	6	6	6	6
	E	2	2	2	2	4	3	2	3	4	2
	F	3	5	5	4	2	2	4	2	2	3
	G	5	4	3	5	3	4	3	5	5	4
	H	7	7	7	6	7	6	7	8	7	7
Dustpan	A	1	1	1	1	1	1	1	1	1	1
	B	4	4	4	4	5	4	4	5	5	5
	C	2	3	3	3	3	3	3	3	2	3
	D	8	8	6	8	7	8	8	8	8	8
	E	6	5	7	7	6	6	6	6	6	6
	F	7	7	8	6	8	7	7	7	7	7
	G	5	6	5	5	4	5	5	4	4	4
	H	3	2	2	2	2	2	2	2	3	2
Flashlight	A	1	1	1	1	1	1	1	1	1	1
	B	6	6	4	4	3	6	3	4	7	6
	C	3	4	7	6	6	4	5	6	3	3
	D	4	5	6	7	7	8	7	7	4	5
	E	8	8	8	8	8	7	8	8	8	8
	F	5	3	5	3	5	5	6	3	5	4
	G	7	7	3	5	4	3	2	5	6	7
	H	2	2	2	2	2	2	4	2	2	2
Frying Pan	A	1	1	1	1	1	1	1	1	1	1
	B	3	4	3	2	6	4	4	3	6	5
	C	2	3	2	6	2	2	2	2	2	3
	D	6	7	6	7	3	6	6	6	3	7
	E	7	8	7	8	4	7	7	7	7	8
	F	8	6	8	5	8	8	8	8	8	6
	G	4	5	5	3	7	5	3	4	4	4
	H	5	2	4	4	5	3	5	5	5	2
Funnel	A	1	1	1	1	1	1	1	1	1	1
	B	7	3	4	3	2	2	4	3	4	2
	C	2	2	2	2	4	3	2	2	2	3
	D	4	7	7	7	8	7	7	7	8	8
	E	8	8	8	8	7	8	8	8	7	7
	F	6	5	6	4	6	5	6	5	6	5
	G	3	6	3	6	3	6	3	6	3	6
	H	5	4	5	5	5	4	5	4	5	4

Hanger	A	1	1	1	1	1	1	1	1	1	1
	B	5	5	5	4	3	4	4	3	6	3
	C	7	8	7	7	7	7	7	7	8	7
	D	2	7	4	5	6	5	3	2	3	4
	E	8	2	8	8	8	8	8	8	7	8
	F	6	3	6	6	4	6	6	6	4	6
	G	4	6	3	3	5	3	5	4	5	5
	H	3	4	2	2	2	2	2	5	2	2
Helmet	A	1	1	1	1	1	1	1	1	1	1
	B	8	7	8	8	8	8	8	7	8	8
	C	5	4	5	3	3	4	4	3	4	4
	D	6	5	6	6	6	6	6	4	6	5
	E	4	6	4	5	5	5	5	6	5	6
	F	7	8	7	7	7	7	7	8	7	7
	G	2	2	2	2	2	2	2	2	2	2
	H	3	3	3	4	4	3	3	3	3	3
Hook	A	1	1	1	1	1	1	1	1	1	1
	B	6	8	7	6	7	7	7	7	6	7
	C	8	7	6	7	8	8	8	8	7	8
	D	2	2	2	2	2	2	2	2	2	2
	E	4	3	4	3	3	5	5	3	3	5
	F	5	5	5	4	4	4	3	5	4	3
	G	7	6	8	8	6	6	6	6	8	6
	H	3	4	3	5	5	3	4	4	5	4
Horn	A	1	1	1	1	1	1	1	1	1	1
	B	2	2	3	2	2	2	3	2	2	2
	C	3	3	2	3	3	3	2	3	3	3
	D	5	4	4	5	4	4	5	5	4	4
	E	4	5	5	4	5	5	4	4	5	5
	F	7	7	6	8	7	7	6	6	6	6
	G	6	6	7	6	6	6	8	6	6	6
	H	8	8	8	7	8	8	7	8	8	8
Kettle	A	1	1	1	1	1	1	1	1	1	1
	B	4	3	2	2	5	3	4	2	2	2
	C	7	4	8	6	8	5	6	4	6	4
	D	5	7	7	7	6	7	7	7	7	8
	E	2	5	3	5	3	4	3	5	5	5
	F	8	8	6	8	7	8	8	8	8	7
	G	3	2	4	3	2	2	2	3	4	3
	H	6	6	5	4	4	6	5	6	3	6

Ladel	A	1	1	1	1	1	1	1	1	1	1
	B	2	2	2	2	3	3	2	2	3	2
	C	6	5	7	7	6	7	3	7	7	8
	D	5	6	8	8	8	8	8	8	8	6
	E	8	4	5	5	5	5	5	4	4	5
	F	3	7	6	6	7	6	6	6	5	7
	G	4	8	3	3	2	2	7	5	2	3
	H	7	3	4	4	4	4	4	3	3	4
Lock	A	1	1	1	1	1	1	1	1	1	1
	B	6	6	6	6	6	6	6	6	6	6
	C	4	3	4	4	4	4	4	2	4	4
	D	5	4	5	5	5	5	5	3	5	5
	E	3	5	3	3	3	3	3	5	3	3
	F	7	7	7	7	7	7	7	7	7	7
	G	2	2	2	2	2	2	2	2	2	2
	H	8	8	8	8	8	8	8	8	8	8
Masher	A	1	1	1	1	1	1	1	1	1	1
	B	8	7	5	5	5	7	6	6	5	5
	C	2	2	2	2	2	2	2	2	2	2
	D	6	5	6	6	7	6	4	5	7	7
	E	4	8	8	8	8	8	8	8	8	8
	F	3	3	3	3	3	4	3	3	3	3
	G	5	4	4	4	4	3	5	4	4	4
	H	7	6	7	7	6	7	7	7	6	6
Necklace	A	1	1	1	1	1	1	1	1	1	1
	B	6	5	4	6	6	5	4	4	7	8
	C	3	7	7	5	5	6	6	6	5	5
	D	4	4	6	4	4	4	5	5	4	4
	E	8	6	5	8	7	7	7	7	6	7
	F	2	2	2	2	3	2	2	2	3	2
	G	5	3	3	3	2	3	3	3	2	3
	H	7	8	8	7	8	8	8	8	8	6
Opener	A	1	1	1	1	1	1	1	1	1	1
	B	8	6	5	3	4	3	5	4	3	4
	C	6	3	3	4	3	4	3	3	4	6
	D	7	4	7	5	6	5	6	6	6	3
	E	4	7	4	7	7	7	8	8	7	7
	F	2	2	2	2	2	2	2	2	2	2
	G	3	8	8	8	8	8	7	7	8	8
	H	5	5	6	6	5	6	4	5	5	5

Paddle	A	1	1	1	1	1	1	1	1	1	1
	B	3	4	3	2	4	3	4	6	4	4
	C	6	2	5	4	2	2	2	2	3	2
	D	5	8	8	7	7	7	7	7	7	7
	E	8	7	7	8	8	8	8	8	8	8
	F	2	6	2	5	6	5	6	4	2	6
	G	4	5	4	6	5	6	5	5	6	5
	H	7	3	6	3	3	4	3	3	5	3
Pitcher	A	1	1	1	1	1	1	1	1	1	1
	B	5	3	3	3	4	2	4	4	4	7
	C	7	7	8	7	7	6	5	7	7	5
	D	4	5	5	4	5	5	7	5	5	3
	E	2	4	4	5	2	3	2	2	3	2
	F	8	8	6	8	8	8	8	8	8	8
	G	6	6	7	6	6	7	6	6	6	6
	H	3	2	2	2	3	4	3	3	2	4
Pliers	A	1	1	1	1	1	1	1	1	1	1
	B	2	3	2	2	4	2	2	2	3	4
	C	4	2	4	3	2	4	4	4	2	2
	D	3	4	3	4	3	3	3	3	4	3
	E	8	8	6	8	8	5	8	6	8	5
	F	6	7	8	7	6	7	7	8	6	7
	G	7	5	5	5	7	8	5	5	7	8
	H	5	6	7	6	5	6	6	7	5	6
Ring	A	1	1	1	1	1	1	1	1	1	1
	B	2	2	2	2	2	2	2	2	2	2
	C	6	3	3	3	3	3	3	3	3	3
	D	8	4	5	4	4	4	4	4	4	4
	E	4	6	6	6	6	6	5	6	6	6
	F	7	5	4	5	5	5	6	5	5	5
	G	3	7	7	7	7	7	7	7	7	7
	H	5	8	8	8	8	8	8	8	8	8
Ruler	A	1	1	1	1	1	1	1	1	1	1
	B	6	5	5	4	6	5	5	6	5	6
	C	8	8	8	8	8	8	8	7	8	7
	D	5	7	6	3	5	6	7	8	7	5
	E	4	4	4	5	4	4	4	4	4	4
	F	2	2	2	2	2	2	2	2	2	2
	G	7	6	7	7	7	7	6	5	6	8
	H	3	3	3	6	3	3	3	3	3	3

Scoop	A	1	1	1	1	1	1	1	1	1	1
	B	2	5	6	6	6	4	6	4	4	4
	C	6	6	4	5	5	5	4	6	5	5
	D	4	2	2	2	2	3	2	2	2	2
	E	3	3	3	3	3	2	3	3	3	3
	F	7	7	5	4	4	6	5	5	6	6
	G	4	4	7	8	8	7	7	7	7	7
	H	8	8	8	7	7	8	8	8	8	8
Screw Driver	A	1	1	1	1	1	1	1	1	1	1
	B	3	5	6	6	3	3	5	5	3	5
	C	6	8	8	8	8	8	8	8	8	8
	D	5	3	3	3	6	4	7	3	6	3
	E	2	4	2	2	2	2	4	2	2	2
	F	7	7	7	7	4	6	2	7	4	6
	G	4	6	5	5	5	5	6	6	5	7
	H	6	2	4	4	7	7	3	4	7	4
Slipper	A	1	1	1	1	1	1	1	1	1	1
	B	3	2	2	3	2	2	3	2	3	2
	C	6	7	7	6	6	7	7	6	8	7
	D	2	3	3	4	3	3	2	3	2	3
	E	4	5	5	5	5	5	4	4	5	5
	F	8	8	8	8	7	8	8	8	7	8
	G	7	6	6	7	8	6	6	7	6	6
	H	5	4	4	2	4	4	5	5	4	4
Spatula	A	1	1	1	1	1	1	1	1	1	1
	B	3	3	3	2	3	2	3	3	2	3
	C	6	5	5	7	5	6	5	6	6	8
	D	4	6	6	3	6	3	6	4	5	6
	E	2	2	2	5	2	4	2	2	4	2
	F	5	4	4	6	4	5	4	5	3	4
	G	8	7	7	4	7	7	7	7	7	5
	H	7	8	8	8	8	8	8	8	8	7
Stapler	A	1	1	1	1	1	1	1	1	1	1
	B	4	4	4	4	5	4	5	4	3	4
	C	2	2	3	2	2	2	2	2	2	2
	D	6	5	2	6	3	7	3	5	5	3
	E	7	7	7	7	6	6	6	7	6	6
	F	3	3	5	3	4	3	4	6	4	5
	G	5	6	6	5	7	5	7	8	7	7
	H	8	8	8	8	8	8	8	3	8	8

Strainer	A	1	1	1	1	1	1	1	1	1	1
	B	3	3	3	2	3	3	3	6	3	3
	C	5	6	5	4	5	4	6	3	5	6
	D	4	5	4	5	4	5	4	4	4	5
	E	6	4	6	6	6	6	5	5	6	7
	F	7	7	8	8	8	8	8	8	8	4
	G	2	2	2	3	2	2	2	2	2	2
	H	8	8	7	7	7	7	7	7	7	8
Thread	A	1	1	1	1	1	1	1	1	1	1
	B	3	5	3	6	5	6	4	5	7	4
	C	6	8	8	8	8	8	8	8	8	8
	D	8	6	4	4	6	4	7	6	5	5
	E	4	3	6	3	3	3	3	3	3	3
	F	2	2	2	2	2	2	2	2	2	2
	G	5	4	5	7	4	7	5	4	6	6
	H	7	7	7	5	7	5	6	7	4	7
Tongs	A	1	1	1	1	1	1	1	1	1	1
	B	2	2	2	2	2	2	2	2	2	2
	C	5	8	5	4	5	4	4	4	5	5
	D	3	3	3	3	4	3	3	3	3	3
	E	4	7	4	5	3	5	8	5	6	4
	F	8	4	6	8	8	8	7	8	8	8
	G	6	5	7	6	6	7	5	6	7	6
	H	7	6	8	7	7	6	6	7	4	7
Whistle	A	1	1	1	1	1	1	1	1	1	1
	B	3	4	4	4	4	3	3	4	3	4
	C	7	7	8	7	7	7	8	8	7	7
	D	6	5	6	5	5	6	6	5	6	5
	E	5	6	5	6	6	5	5	6	5	6
	F	4	3	3	3	3	4	4	3	4	3
	G	2	2	2	2	2	2	2	2	2	2
	H	8	8	7	8	8	8	7	7	8	8
Wrench	A	1	1	1	1	1	1	1	1	1	1
	B	3	8	8	8	8	8	8	8	3	8
	C	7	4	4	6	5	6	5	7	7	4
	D	8	7	3	3	7	7	7	3	8	7
	E	5	6	7	7	4	3	3	4	6	6
	F	6	5	5	4	6	5	6	6	4	5
	G	2	2	2	2	2	2	2	2	2	2
	H	4	3	6	5	3	4	4	5	5	3

Zipper

A	1	1	1	1	1	1	1	1	1	1
B	2	3	2	2	3	2	4	2	2	2
C	8	8	8	8	8	8	8	8	8	8
D	6	6	7	6	7	6	6	6	6	6
E	7	7	6	7	6	7	7	7	7	7
F	5	4	4	4	4	4	5	5	5	4
G	3	2	3	3	2	3	3	3	3	3
H	4	5	5	5	5	5	2	2	4	5

Appendix C

Best Example Theory Rating Instructions

The instructions used for rating objects according to the best example theory were identical to those used by Hupp (1980) and were as follows:

This study has to do with what we have in mind when we use words which refer to categories. Take the word "red" as an example. Close your eyes and imagine a true red. Now imagine an orange-red or a purplish-red. Although you might still name the orange-red or the purple-red with the "red", they are not as good examples of red (as clear cases of what "red" refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other types of categories. For example, think of fruits. An apple is a much better example of "fruit" than a coconut is. Note that this kind of judgement has nothing to do with how well you like the thing. You might like a purple-red better than a true red, but still recognize that the color you like is not a true red. You may prefer to eat coconuts than to eat apples, without thinking that coconut better represents what people mean by "fruit".

On this form you are asked to judge how good an example of a category various objects are. First look at all of the objects in the category. You will notice that each of the objects has a letter on it. On the form, the object letters are listed.

After each letter there is a blank. Please rate how good an example of the category each object is, using a 7-point scale. A "1" means that you feel the object is a very good example of your idea or image of the category. A "7" means you feel that the object fits very poorly with your idea or image of the category. A "4" means you feel the object fits moderately well. Using the other numbers of the 7-point scale to indicate intermediate judgements.

While all the objects have been considered members of the category by at least some people, you may disagree. If you feel that an object is not a member of the category, put an "X" in the blank after its number.

Do not worry about why you feel that something is or is not a good example of the category. And do not worry about whether it is just you, or whether it is people in general, who feel that way. Just mark it the way you see it. You may rate the objects in any order that you wish.

Appendix D

Results of the Best Example Ratings For Experiment 1

The results from the best example ratings were as follows:

Stimulus Class	Stimulus Members							
	A	B	C	D	E	F	G	H
Badge $\bar{x} =$	1.1	1.1	1.2	3.7	3.8	1.2	2.0	1.4
Beater $\bar{x} =$	1.9	1.7	5.0	2.6	2.1	1.4	1.6	5.0
Bracelet $\bar{x} =$	1.9	2.4	3.6	1.2	2.0	3.2	2.0	1.6
Bowl $\bar{x} =$	1.3	1.4	2.0	1.3	1.5	2.2	2.2	2.3
Candle $\bar{x} =$	1.1	1.4	1.2	1.0	1.8	2.3	1.4	2.1
Chain $\bar{x} =$	1.2	1.5	3.0	2.5	1.8	1.2	1.9	1.9
Clothespin $\bar{x} =$	1.2	3.8	4.2	1.9	4.4	4.1	4.1	2.2
Clamp $\bar{x} =$	2.1	4.1	1.7	2.0	2.1	1.0	1.2	2.2
Clip $\bar{x} =$	1.5	2.9	3.1	2.5	1.3	3.5	3.6	4.6
Coffeepot $\bar{x} =$	2.0	1.6	4.4	3.9	2.0	1.7	1.9	4.3
Dustpan $\bar{x} =$	1.3	1.5	1.5	3.5	2.0	4.7	1.6	1.2
Flashlight $\bar{x} =$	1.3	1.4	1.5	1.7	3.9	1.5	1.3	1.5
Frying Pan $\bar{x} =$	1.0	1.6	1.6	1.7	1.4	4.2	2.1	1.0
Funnel $\bar{x} =$	1.0	1.6	1.0	2.1	3.0	2.1	1.1	1.4

Stimulus Class	Stimulus Members							
	A	B	C	D	E	F	G	H
Hanger								
$\bar{x} =$	1.5	1.6	1.2	2.0	1.3	2.2	1.5	1.9
Helmet								
$\bar{x} =$	1.1	3.1	1.1	3.3	4.2	4.4	1.2	1.3
Hook								
$\bar{x} =$	2.1	1.8	3.1	2.4	1.6	2.1	3.9	2.3
Horn								
$\bar{x} =$	1.0	3.1	3.3	2.2	2.3	2.6	3.4	4.1
Kettle								
$\bar{x} =$	1.5	1.6	2.0	2.6	1.6	2.4	1.2	2.9
Ladel								
$\bar{x} =$	1.4	1.6	2.9	3.1	4.0	2.0	1.3	3.8
Lock								
$\bar{x} =$	1.0	1.0	1.1	1.1	5.3	1.8	1.4	1.9
Masher								
$\bar{x} =$	1.8	3.2	2.5	2.0	4.2	2.5	2.9	3.5
Necklace								
$\bar{x} =$	1.3	2.5	1.8	1.7	2.0	1.8	1.9	2.6
Opener								
$\bar{x} =$	1.5	1.3	2.2	2.8	4.2	1.4	1.8	3.8
Paddle								
$\bar{x} =$	2.4	3.0	2.2	4.2	5.1	3.0	2.4	2.6
Pitcher								
$\bar{x} =$	1.1	2.0	2.4	2.2	1.3	3.6	2.6	1.6
Pliers								
$\bar{x} =$	1.2	1.3	1.3	1.3	2.8	2.4	4.6	2.1
Purse								
$\bar{x} =$	1.2	3.9	5.0	2.6	2.4	5.5	2.7	3.2
Ring								
$\bar{x} =$	1.2	1.5	1.3	1.9	1.5	2.1	1.4	2.3
Ruler								
$\bar{x} =$	1.2	2.3	3.3	2.4	1.6	1.8	3.0	1.1

Stimulus Class	Stimulus Members							
	A	B	C	D	E	F	G	H
Scoop $\bar{x} =$	1.8	2.2	2.2	2.1	2.1	2.0	1.6	2.7
Screwdriver $\bar{x} =$	1.2	1.2	4.6	1.8	1.7	1.5	1.7	2.9
Slipper $\bar{x} =$	1.5	1.8	3.7	2.5	2.5	2.2	1.1	4.3
Spatula $\bar{x} =$	1.6	1.8	1.7	1.4	2.0	2.5	1.5	3.3
Stapler $\bar{x} =$	1.1	1.5	1.5	1.8	2.2	1.9	2.2	3.5
Strainer $\bar{x} =$	1.1	2.2	1.1	1.1	1.9	3.9	1.6	2.7
Thread $\bar{x} =$	1.0	1.1	1.6	1.1	1.0	1.0	1.0	1.1
Tongs $\bar{x} =$	1.9	1.4	2.9	1.4	2.7	4.1	2.3	3.4
Whistle $\bar{x} =$	1.0	1.6	2.9	2.4	2.8	1.5	1.1	4.2
Wrench $\bar{x} =$	1.4	2.3	2.8	4.0	2.6	4.3	1.1	2.3
Zipper $\bar{x} =$	1.9	2.7	1.4	2.6	2.5	1.9	1.7	1.8

Appendix E

Group Rating Instructions For Experiment 2

The instructions used for determining the two different groups for each stimulus class were as follows:

On the table in front of you are many different groups of objects. Each group is commonly referred to by the same name. Thus, as you can see there is a group containing all hammers and another containing all kettles and so forth. Within each group there are 8 objects, some very similar to one another, others quite different.

For each group, I would like you to place them into one of two categories, either 1 or 2, using similarity as a basis for categorizing them. An object placed in one category should ideally share many characteristics with the other objects in that category and share fewer characteristics with the objects in the other category.

If you cannot place an object into either Category 1 or 2, or you believe that it does not belong to that group of objects set it aside. There is no set number of objects that you should place in Category 1 or 2. Once you have placed the objects into a category please write down on the sheet in front of you, under the appropriate category the letter that is printed on each object.

I would like you to remember that this is not a test. Different people judge things in different ways and there are no right or wrong answers. Two objects that are very similar to one person may be quite different to another. Both results are important to me. I am interested in how you as an individual rate these objects.

Appendix F

Forced Choice Rating Instructions For Experiment 2

The instructions for the forced choice rating were as follows:

On the table in front of you are many different groups of objects. Each group is commonly referred to by the same name. Thus, as you can see there is a group containing all hammers and another containing all kettles and so forth. Within each group there are 8-14 objects, some very similar to one another, others quite different.

For each group, I would like you to place the objects into one of two categories, either 1 or 2, using similarity as a basis for categorizing them. An object placed in one category should ideally share many characteristics with the other objects in that category and share fewer characteristics with the objects in the other category. You should place 4 objects in each category. If you find that an object cannot be easily placed in either category please place it with the objects that are in your judgement most similar to it, keeping in mind that 4 objects must be placed in each category. Once you have placed the objects into a category please write on the sheet in front of you, under the appropriate category the letter that is printed on each object.

I would like you to remember that this is not a test. Different people judge things in different ways and there are no right or wrong answers. Two objects that are very similar to one person may be quite different to another. Both results are important to me. I am interested in how you as an individual rate these objects. Do you have any questions?

APPENDIX G

Results of the Best Example Ratings For Experiment 2
 The results from the best example ratings were as follows:

Exemplars	Stimulus Members							
	Group 1				Group 2			
Badge	H	F	B	C	J	D	E	I
\bar{x} =	1.4	1.1	1.1	1.2	3.9	3.7	3.8	3.3
Beater	A	J	B	I	G	D	F	K
\bar{x} =	1.8	1.5	1.7	2.4	1.6	2.6	1.4	2.1
Bracelet	E	C	M	O	I	J	K	L
\bar{x} =	2.0	3.7	4.3	1.7	1.6	2.0	2.6	1.8
Brush	C	I	E	L	G	A	H	K
\bar{x} =	2.5	3.7	4.3	1.4	3.8	3.7	3.5	3.8
Bowl	B	A	I	J	G	F	H	K
\bar{x} =	1.8	2.1	2.2	1.8	1.4	1.3	1.3	1.5
Candle	C	F	D	I	E	A	M	N
\bar{x} =	1.2	1.1	1.0	1.1	1.8	2.3	1.8	1.7
Chain	F	B	A	J	L	C	K	M
\bar{x} =	1.2	1.5	1.2	1.7	3.1	3.0	2.7	3.5

Clothespin	K	S	D	R	N	E	O	P
$\bar{x} =$	1.0	2.0	1.9	2.4	3.4	4.4	2.8	4.6
Coffeepot	F	B	A	G	H	C	D	I
$\bar{x} =$	1.7	1.6	2.0	1.9	4.3	4.4	3.9	3.8
Comb	E	L	I	K	F	G	H	A
$\bar{x} =$	1.3	1.3	1.3	1.3	1.4	1.4	1.7	1.3
Dustpan	A	C	H	J	G	B	E	I
$\bar{x} =$	1.3	1.5	1.2	1.3	1.5	2.0	1.5	1.4
Flashlight	B	J	H	I	D	A	E	F
$\bar{x} =$	1.3	1.8	1.5	1.8	1.7	1.4	3.9	1.5
Frying Pan	D	G	B	C	I	F	J	K
$\bar{x} =$	1.0	2.1	1.6	1.7	5.3	4.2	5.3	5.4
Funnel	A	N	B	H	J	I	K	L
$\bar{x} =$	1.0	1.3	1.6	1.4	4.3	3.9	5.1	4.3
Hammer	G	E	B	I	F	C	D	H
$\bar{x} =$	1.0	1.6	1.0	3.2	4.5	5.2	4.6	4.4
Hanger	D	H	F	G	L	E	I	M
$\bar{x} =$	1.5	1.2	1.8	2.4	3.1	3.5	3.3	3.2

Helmet	C	H	G	J	I	B	F	K
$\bar{x} =$	1.1	1.3	1.2	1.1	2.5	3.1	4.6	1.5
Hook	E	F	A	H	K	J	L	M
$\bar{x} =$	1.7	2.1	2.1	2.3	1.6	2.1	1.5	1.4
Horn	G	E	D	J	C	B	H	L
$\bar{x} =$	1.0	2.3	2.2	1.4	3.3	3.1	4.1	3.3
Kettle	A	G	B	E	O	C	F	I
$\bar{x} =$	1.6	1.2	1.6	1.6	2.6	2.0	2.4	2.5
Ladel	I	L	G	J	H	A	C	E
$\bar{x} =$	1.4	2.2	1.3	1.6	3.8	1.3	2.9	4.0
Lock	F	C	D	G	A	B	H	J
$\bar{x} =$	1.0	1.1	1.1	1.4	1.7	1.0	1.9	1.9
Masher	A	C	F	I	D	H	L	M
$\bar{x} =$	1.8	2.5	2.5	4.0	1.9	3.5	2.4	2.2
Necklace	I	C	E	G	A	J	L	M
$\bar{x} =$	1.3	1.8	2.0	1.9	1.4	2.0	1.6	1.6
Opener	D	E	C	H	L	I	K	G
$\bar{x} =$	1.5	4.2	2.2	3.8	3.0	3.0	2.9	2.0

Pitcher	I	M	O	P	E	A	D	H
$\bar{x} =$	1.6	1.7	1.8	3.1	1.3	1.1	1.3	1.6
Pliers	A	C	B	I	H	F	K	L
$\bar{x} =$	1.2	1.3	1.3	5.0	2.1	2.4	4.0	4.1
Purse	G	E	A	D	B	F	J	K
$\bar{x} =$	1.9	1.7	1.0	2.4	3.9	4.6	4.4	4.4
Ring	G	I	H	O	B	L	M	N
$\bar{x} =$	1.4	2.3	2.3	2.7	1.5	1.7	2.5	2.7
Ruler	B	F	H	I	J	A	G	L
$\bar{x} =$	1.2	1.8	1.1	1.3	2.6	2.3	3.0	2.5
Scoop	F	E	D	I	G	A	H	J
$\bar{x} =$	1.8	2.1	2.1	2.7	1.6	2.0	2.7	1.6
Screwdriver	B	F	E	G	J	M	N	O
$\bar{x} =$	1.2	1.5	1.7	1.7	3.5	4.2	2.7	2.7
Slipper	B	I	A	F	E	C	H	J
$\bar{x} =$	1.8	1.2	1.5	2.2	2.5	3.7	4.3	1.6
Spatula	E	M	I	L	D	C	G	H
$\bar{x} =$	2.0	2.7	2.8	2.4	1.4	1.7	1.5	3.3

Stapler	A	J	C	K	G	E	F	K
$\bar{x} =$	1.1	1.4	1.5	1.0	2.2	2.2	1.9	2.2
Strainer	B	E	C	G	I	J	K	L
$\bar{x} =$	2.2	1.9	1.1	1.6	2.8	2.2	2.4	2.3
Thread	C	H	E	I	A	J	L	M
$\bar{x} =$	1.0	1.1	1.1	1.2	1.6	1.9	2.2	2.3
Tongs	H	D	B	K	I	F	A	J
$\bar{x} =$	1.9	1.4	1.4	2.1	3.7	4.1	3.4	3.1
Whistle	C	J	G	I	D	A	O	P
$\bar{x} =$	1.0	1.2	1.1	1.0	2.4	2.9	2.8	2.9
Wrench	O	K	I	J	H	C	F	L
$\bar{x} =$	1.7	4.6	1.5	2.0	2.3	2.8	4.3	2.3

Appendix H

Similarity Rating Instructions For Experiment 2

The instructions used in Experiment 2 for rating each stimulus member of a stimulus class. Each member was rated to one of the training stimuli using similarity as a basis for rating them. The following instructions were used:

On the table in front of you are many different groups of objects. Each group is commonly referred to by the same name. Thus, as you can see there is a group containing all hammers, another containing all kettles and so forth. Within each group there are 8 objects.

Each object within a group has been labelled with different letters. I would like you to first locate the object with the same letter as that found beside the object name on your data sheet (the target object). I would then like you to study for a moment the remaining 7 objects in relation to the target object. After studying them I would like you to rate the objects with the remaining letters, one at a time, in relation to the target object using similarity as a basis for rating them. You are to make your ratings using the 7 point scale in front of you.

If you find that one object of a group is very similar to the target object of that group you should mark the first point on the scale above the descriptor labelled "Very Similar". If you find that an object is completely different from the target object you should mark the seventh point on the scale above the descriptor labelled "Completely Different". If you find that the object falls somewhere between these two end points you should mark off the space that best represents how similar that object is to the target object.

I would like you to remember that this is not a test. Different people judge things in different ways and there are no right or wrong answers. Two objects that are very similar to one person may be quite different to another. Both results are important to me. I am interested in how you as an individual rate these objects.

Appendix I

Shape Rating Instructions For Experiment 2

The instructions used in Experiment 2 for rating each stimulus member of a stimulus class. Each member was rated to one of the training stimuli using shape as a basis for rating them. The following instructions were used:

On the table in front of you are many different groups of objects. Each group is commonly referred to by the same name. Thus, as you can see there is a group containing all hammers, another containing all kettles and so forth. Within each group there are 8 objects.

Each object within a group has been labelled with different letters. I would like you to first locate the object with the same letter as that found beside the object name on your data sheet (the target object). I would then like you to study for a moment the remaining 7 objects in relation to the target object. After studying them I would like you to rate the objects with the remaining letters, one at a time, in relation to the target object using shape as a basis for rating them. You are to make your ratings using the 7 point scale in front of you.

If you find that one object of a group is very similar in shape to the target object of that group you should mark the first point on the scale above the descriptor labelled "Very Similar". If you find that an object is completely different in shape from the target object you should mark the seventh point on the scale above the descriptor labelled "Completely Different". If you find that the object falls somewhere between these two end points you should mark off the space that best represents how similar in shape that object is to the target object.

I would like you to remember that this is not a test. Different people judge things in different ways and there are no right or wrong answers. Two objects that are very similar in shape to one person may be quite different to another. Both results are important to me. I am interested in how you as an individual rate these objects.

Appendix J

The Results of the Similarity (SIM) and Shape (SH)
Ratings For Experiment 2

The data for the similarity and shape ratings are presented as the mean rating. Each letter indicates a training or testing stimulus. The blank spaces indicate the potential training object that was rated in relation to the remaining seven objects. Only the data for the stimulus classes that were trained are presented.

Stimulus
Members

Badge	SIM	SH	SIM	SH	SIM	SH
Group 1						
H			2.7	2.1	6.2	5.0
F	2.7	2.7			6.8	5.5
B	2.2	2.2	2.5	2.1	6.2	5.2
C	1.6	2.1	2.8	2.2	6.3	5.6
Group 2						
J	5.4	4.9	5.5	5.4		
D	5.9	6.0	6.3	6.6	2.9	3.9
E	5.6	5.6	6.0	6.5	2.9	3.4
I	5.7	6.1	6.7	5.6	3.2	4.6

Stimulus
Members

Beater	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			2.6	2.8	4.8	4.0
J	2.2	2.7			4.8	4.2
B	1.7	2.0	2.0	1.8	4.6	3.4
I	3.3	3.7	3.6	3.2	5.6	4.2
Group 2						
G	3.7	4.0	3.9	4.0		
D	4.6	4.4	5.1	4.3	3.7	3.2
F	3.2	3.9	3.4	4.1	1.4	1.3
K	3.4	3.8	4.2	4.6	4.0	4.1
Bracelet						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
E			4.0	4.0	5.6	5.3
C	3.2	4.0			6.3	4.8
M	3.2	3.7	3.7	3.7	6.7	5.1
O	1.8	2.4	3.8	3.4	4.7	4.9
Group 2						
I	4.9	5.1	5.3	5.9		
J	5.2	5.4	6.1	6.2	2.8	3.0
K	4.9	5.7	5.9	6.5	3.8	4.3
L	5.1	5.0	6.1	5.9	3.3	2.3

Stimulus
Members

Chain	SIM	SH	SIM	SH	SIM	SH
Group 1						
B			2.3	3.0	5.9	4.7
E	3.2	2.4			6.7	5.3
C	3.5	3.0	1.7	2.0	6.6	5.5
H	1.8	1.3	3.0	3.5	6.0	4.5
Group 2						
L	4.3	4.4	5.6	6.0		
G	5.3	4.4	5.6	5.8	3.9	3.5
I	4.7	3.7	5.3	5.6	2.5	2.9
K	5.3	5.0	5.9	6.1	2.6	3.1
Clothespin						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
K			3.4	3.2	6.1	4.8
S	3.1	2.7			5.2	4.4
D	1.8	1.8	2.3	2.5	4.7	4.2
R	3.1	2.2	1.1	1.0	5.1	4.4
Group 2						
N	4.0	5.6	4.3	5.4		
E	5.2	5.2	4.7	5.6	3.1	3.4
O	4.1	4.4	4.2	5.3	2.5	2.1
P	4.3	6.2	5.2	5.9	3.7	3.5

Stimulus
Members

Coffeepot	SIM	SH	SIM	SH	SIM	SH
Group 1						
B			2.5	3.7	5.9	4.5
F	2.5	2.8			5.4	5.3
A	2.3	2.1	1.5	1.4	6.2	4.3
G	3.9	3.7	2.5	3.4	6.2	4.6
Group 2						
H	5.6	5.4	5.9	5.7		
C	5.4	5.7	6.6	6.3	3.3	3.6
D	5.8	5.3	4.9	5.5	3.7	3.5
I	6.2	5.5	6.4	6.2	3.1	3.6
Dustpan						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			1.3	1.8	6.0	4.2
C	1.6	1.4			5.7	4.5
H	1.3	1.3	1.2	1.7	5.5	4.3
J	3.0	2.9	2.8	2.5	5.8	5.2
Group 2						
G	4.6	3.6	5.1	5.1		
B	4.7	3.4	4.7	5.1	2.0	2.2
E	5.2	4.9	5.4	6.2	3.6	3.3
I	4.3	3.9	5.0	4.8	1.8	1.5

Stimulus
Members

Funnel	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			3.2	2.9	5.9	4.8
N	3.1	3.2			6.2	5.0
B	1.6	1.8	2.6	2.7	6.3	5.3
H	2.8	2.2	3.5	3.7	5.6	5.0
Group 2						
J	5.7	6.4	5.2	6.0		
I	5.9	6.2	5.4	6.1	2.6	2.9
K	6.4	6.4	5.7	6.4	3.2	2.5
L	5.6	6.2	5.0	6.1	1.4	1.3
Helmet						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
C			4.1	4.1	6.6	5.2
H	4.1	3.6			6.2	4.8
G	3.8	3.1	3.7	3.6	6.5	5.6
J	2.0	2.0	5.1	4.0	6.5	5.9
Group 2						
I	5.4	5.9	5.1	6.0		
B	6.4	6.0	6.2	6.4	4.3	3.8
F	6.2	6.1	5.9	6.3	4.1	3.0
K	5.3	5.0	4.9	5.6	3.5	3.1

Stimulus
Members

Hook	SIM	SH	SIM	SH	SIM	SH
Group 1						
F			2.4	2.5	5.3	5.3
E	2.5	2.3			6.2	5.3
A	2.5	2.8	2.4	3.0	6.3	5.6
H	1.9	2.2	3.4	3.0	5.7	4.8
Group 2						
J	4.8	5.2	6.1	6.3		
K	5.1	5.4	6.0	6.6	3.0	1.6
L	4.8	5.6	5.8	5.9	2.5	1.9
M	6.0	6.7	6.5	6.9	5.6	4.3
Kettle						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			2.6	2.1	5.0	4.8
G	2.6	2.8			4.5	5.0
B	2.1	2.2	3.7	3.7	5.0	4.7
E	2.8	3.1	1.9	2.6	5.4	5.2
Group 2						
D	4.5	5.1	5.8	5.8		
C	4.1	4.9	5.8	5.5	3.3	2.6
F	4.5	5.8	6.0	6.4	2.3	3.1
I	4.7	5.8	5.6	6.1	2.4	1.8

Stimulus
Members

Ladel	SIM	SH	SIM	SH	SIM	SH
Group 1						
L			2.2	2.9	5.5	4.6
I	2.4	2.7			6.0	4.7
G	2.4	3.2	2.8	3.3	5.7	4.9
J	2.9	3.4	1.9	2.7	6.4	4.9
Group 2						
H	5.6	3.7	5.8	5.5		
A	4.9	5.0	5.1	5.1	4.6	3.3
C	5.5	5.2	5.6	6.3	2.9	3.5
E	5.4	3.7	5.9	5.2	1.5	1.1
Lock						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
F			3.2	3.1	5.8	5.3
C	3.1	2.7			6.4	5.6
D	3.7	3.8	1.1	1.7	6.0	5.4
G	1.6	2.0	3.4	2.9	5.9	5.5
Group 2						
A	5.9	6.0	6.1	6.4		
B	4.6	2.0	5.3	4.9	3.6	3.2
H	6.6	5.4	6.1	6.4	3.5	2.8
J	5.8	6.0	6.2	6.2	1.4	1.1

Stimulus
Members

Masher	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			1.9	2.4	5.5	4.8
C	2.2	1.6			5.5	5.5
F	3.5	3.0	2.7	3.6	4.6	5.3
I	3.6	3.3	3.7	4.5	6.1	5.5
Group 2						
D	5.6	5.4	5.2	5.5		
H	5.0	5.2	6.0	5.9	4.0	4.4
L	5.6	5.4	5.5	5.4	2.7	1.6
M	5.6	5.4	5.5	5.1	2.0	1.2
Opener						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
D			3.7	5.6	5.5	4.7
E	4.1	4.8			6.6	5.7
C	3.2	4.3	4.1	4.4	5.3	5.0
H	3.8	4.9	3.4	4.3	6.1	5.1
Group 2						
L	5.1	5.1	5.9	6.1		
I	4.9	5.3	6.0	6.4	3.7	2.5
G	5.0	5.2	6.3	6.4	1.4	1.1
K	5.1	5.2	5.6	6.5	3.9	2.9

Stimulus
Members

Pitcher	SIM	SH	SIM	SH	SIM	SH
Group 1						
I			1.1	1.0	5.8	5.1
M	1.2	1.1			5.7	5.0
O	2.8	2.6	2.5	2.8	6.1	5.1
P	4.8	5.6	4.8	6.4	6.3	5.7
Group 2						
E	5.2	4.9	4.8	5.4		
A	5.3	5.0	4.9	5.1	1.3	1.5
D	6.0	6.1	5.8	6.4	3.8	3.8
H	4.9	5.6	5.3	6.2	4.1	3.7
Pliers						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			1.6	1.4	4.9	4.9
C	1.4	1.3			6.0	5.5
B	1.5	1.3	1.8	1.8	5.3	5.0
I	4.4	2.4	4.3	2.7	6.8	5.8
Group 2						
H	5.0	4.9	5.2	5.5		
F	5.1	4.4	5.4	5.7	1.8	1.5
K	5.8	5.2	6.2	6.3	3.6	3.0
L	5.6	5.6	6.1	6.2	4.5	3.3

Stimulus
Members

Ruler	SIM	SH	SIM	SH	SIM	SHG
-------	-----	----	-----	----	-----	-----

Group 1

B			3.6	3.0	6.2	4.8
F	3.2	2.3			5.9	4.7
H	3.1	2.8	3.3	3.8	5.5	3.8
I	3.5	2.4	3.6	3.8	5.6	4.1

Group 2

J	5.1	4.2	5.7	4.2		
A	5.5	3.9	6.0	5.3	3.6	3.0
G	5.4	4.6	6.2	5.0	2.8	2.3
L	5.6	4.4	6.0	5.1	4.0	3.1

Scoop	SIM	SH	SIM	SH	SIM	SH
-------	-----	----	-----	----	-----	----

Group 1

F			2.5	2.6	4.5	4.6
E	2.6	2.4			5.6	5.0
D	1.5	1.3	2.9	3.4	5.6	4.8
I	2.7	3.0	2.2	3.2	5.9	5.1

Group 2

G	4.3	4.9	5.7	5.6		
A	4.3	4.4	4.5	4.6	4.4	4.9
H	5.5	5.8	5.6	6.2	2.8	3.6
J	4.1	4.9	5.6	5.3	1.1	1.1

Stimulus
Members

Screwdriver	SIM	SH	SIM	SH	SIM	SH
Group 1						
B			3.6	4.6	5.8	4.7
F	3.6	4.0			5.1	4.4
E	3.1	2.1	2.2	4.0	4.8	4.1
G	2.6	3.7	3.5	2.8	5.5	3.9
Group 2						
J	5.4	4.0	5.0	5.1		
M	5.7	6.4	5.5	6.4	3.8	3.9
N	5.0	4.3	4.9	5.3	1.8	2.3
O	5.7	4.6	6.2	5.9	3.0	2.4
Spatula						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
E			3.1	2.9	6.0	5.3
M	3.2	3.3			5.2	4.3
I	1.4	1.6	3.1	3.1	5.6	5.2
L	3.5	3.6	3.2	2.8	5.4	5.3
Group 2						
D	4.8	5.0	4.1	3.9		
C	4.3	3.8	4.7	3.9	3.0	2.2
G	4.9	5.3	5.1	4.3	1.8	2.4
H	5.0	5.8	5.7	4.9	4.9	3.6

Stimulus
Members

Stapler	SIM	SH	SIM	SH	SIM	SH
Group 1						
A			1.3	1.5	5.2	4.6
J	1.4	1.5			5.6	4.6
C	2.7	3.0	2.4	3.2	5.7	4.7
K	2.5	2.7	2.4	3.3	5.2	5.1
Group 2						
G	4.2	4.2	5.2	5.1		
E	4.4	4.7	5.1	4.5	3.0	2.4
F	3.9	3.4	4.2	3.9	2.3	2.5
I	4.3	4.1	5.1	4.4	2.4	1.7
Strainer						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
B			2.1	2.6	6.6	5.5
E	2.7	2.0			5.9	5.4
C	2.5	2.1	1.3	2.4	5.1	4.8
G	1.8	2.2	2.8	3.4	5.2	5.2
Group 2						
I	6.2	6.4	6.1	6.4		
J	5.3	5.2	5.0	5.7	3.8	4.6
K	6.1	6.4	6.4	6.7	2.5	3.0
L	5.9	6.0	5.0	5.8	2.0	1.8

Stimulus
Members

Thread	SIM	SH	SIM	SH	SIM	SH
Group 1						
C			5.0	2.9	6.2	5.6
H	3.0	2.4			5.8	5.0
E	2.2	1.9	3.4	1.9	6.1	4.8
I	2.8	2.3	2.5	1.5	5.8	4.9
Group 2						
A	5.0	5.6	5.6	5.4		
J	4.9	5.5	5.7	5.0	2.5	1.2
L	4.9	5.7	6.1	5.8	3.1	2.3
M	5.2	5.6	6.0	6.1	2.7	2.5
Tongs						
	SIM	SH	SIM	SH	SIM	SH
Group 1						
H			2.9	3.3	6.5	6.3
D	2.9	3.2			6.7	5.9
B	2.6	2.6	2.1	2.9	6.3	5.7
K	3.5	3.1	2.2	2.5	6.6	5.7
Group 2						
I	5.8	6.4	6.1	6.6		
F	5.7	6.5	6.4	6.3	2.8	2.9
A	5.5	6.1	6.1	6.8	2.6	2.5
J	5.7	6.3	6.1	6.7	4.2	3.6

Stimulus
Members

Wrench	SIM	SH	SIM	SH	SIM	SH
Group 1						
O			2.3	1.8	6.5	5.8
K	3.2	1.7			6.9	4.9
I	1.2	1.6	2.5	1.7	6.4	6.4
J	1.7	1.8	3.0	2.0	6.3	4.9
Group 2						
H	6.2	6.5	6.6	6.8		
C	6.3	6.4	6.5	6.6	2.4	3.9
F	6.4	6.5	6.8	6.5	2.7	3.8
L	6.3	6.5	6.5	6.7	3.0	4.2