INSTITUTE FOR THE
STUDY OF
INTELLECTUAL
BEHAVIOR

Typicality Effects in Logically-defined Categories[1]

Lyle E. Bourne, Jr.
University of Colorado

Technical Report No. 96
Institute for the Study of Intellectual Behavior
University of Colorado
May 1980

ABSTRACT

Subjects learned to categorize geometric designs according to two features, x and y. Where xy stimuli (stimuli with both features) are always positive, the concept is inclusive disjunctive; where xy stimuli are always negative, the concept is exclusive disjunctive. In three other experimental conditions, the probability of xy stimuli being positive during acquisition was .25, .50 and .75. In a variety of post-acquisition tests, xy stimuli were chosen as prototypical despite their less than consistent occurrence in the positive category. Stimuli with one relevant feature (x or y), though consistently assigned to the positive category, were evaluated as poorer examples of the concept. These results are interpreted in terms of a schema model in which information acquired during learning is organized according to probability density functions over feature dimensions. This theory appears general enough to accomodate the evidence from laboratory studies of both logical and natural concepts.

Conventional laboratory studies of concept learning have been criticized for their simplicity and artificiality, and their failure to tell us anything useful about natural categorization or category organization (Rosch, 1978). These experiments typically use simple stimuli which take on only a small number of discrete values on a small number of separable, orthogonal dimensions (Garner, 1978), with each dimension represented by the same number of values. The stimulus features and the categories to be learned are already well-known by the subject. Every instance of a particular category, for example, "red square," is as good as any other instance of the same category. In contrast to "laboratory" concepts, natural categories appear to be ill-defined and not nearly so clear, or artificial, or internally unstructured. Instances of natural categories vary on a large number not necessarily orthogonal dimensions. Furthermore, each dimension may have many or few and not necessarily the same number of values. In fact, many dimensions are continuous rather than discrete. One instance of a concept is not necessarily as good as any other. Indeed, natural categories often seem to be based on "typical" or "best" instances. According to Rosch (1978) these prototypes form a focus for the category and serve as the basis for organizing our perception of category instances. They can be considered as memory representations of categories in analog format. In a variety of experiments, Rosch and Mervis (1975); Rosch, Mervis, Gray, Johnson and Boyes-Braem, (1976) demonstrated the reality of prototypes within natural categories and a variety of empirical effects attributable to the varying degrees of membership of category instances.

There are, indeed, important limitations on the generality of traditional laboratory experiments in so far as the formation and knowledgeable use of natural categories is concerned. One need not despair of these criticisms, however, for traditional laboratory experiments by and large were not intended to deal directly with the formation and use of natural categories. They focus primarily on the formation of well-defined logical concepts, rather than ill-defined natural or semantic concepts. Further, these traditional studies have attempted primarily to discover the processes by means of which relevant attributes or relationships among relevant attributes are identified, or learned, rather than to examine the structure of these categories or their memory representation. But beyond that, the general theoretical position which has been the basis for traditional work can accommodate phenomena like typicality which have been identified within natural concepts. Thus, the existence of those phenomena in no way vitiates traditional theory or traditional methodology.

Consider the following arguments. Suppose people mentally represent concepts as dimensionalized schemata. The subject enters any new concept learning problem with a generalized schema containing empty slots corresponding to known dimensions of the stimuli to be categorized. The slots act as requests for information about these dimensions from the task (stimuli and feedback). As slots fill with information, some dimensions are identified as relevant, because of their contingency with feedback, and conversely others are identified as irrelevant. At problem solution, the subject has a concept-specific schema which contains pertinent information on relevant and irrelevant stimulus dimensions in the appropriate slots (or could conceivably lack any slots for irrelevant information altogether).

On this view, people become aware of or focus upon stimulus dimensions which exhibit, over a sample of stimuli, some contingency with the positive (or possibly negative or contrast) category of a concept. Commonly-occurring features within the range of possibilities on a given dimension become the focus of attention and enter into a person's understanding and representation of the concept. Thus, a concept can be defined in terms of some relationship among a set of probability functions, one for each identifiable dimension of the stimuli. Under such an assumption, a positive instance of the concept need *not* have the most probable value on each defining dimension of the concept. A bird, for example, is likely to be able to fly, have two legs, a beak, and feathers. But a "footless plucked chicken with its head cut off" is still a bird despite its lack of these high probability features.

This probabilistic characterization of a concept deals, to our knowledge, with all of the phenomena identified in studies of natural concepts. For example, for X to be a defining dimension, $p(X)$, a probability density function, must be "significantly" non-rectangular for the category of positive instances. If it is non-rectangular, then it contains at least one value, x_m with a maximal contingency with the concept. Now the definition of a prototype, at least for conjunctive and inclusive disjunctive concepts, is straightforward. The prototypical stimulus will be the stimulus which contains values x_m , on all defining dimensions, X. Even if the subject has never seen the prototypical stimulus configuration, he will still identify it as the "best example" of the concept category, for this is the modal stimulus in the relevant multi-dimensional similarity space. The degree of category membership is also straightforwardly calibrated. The degree of category membership for any stimulus will be a function of (*a*) how many x_m 's it exhibits, (b) how close its value, x_i , on dimension X is to x_m , in the similarity space, and finally (c) which x_m 's are involved. The latter criterion takes into account the fact that the peak values on the various defining dimensions can themselves differ in probability.

In this theory, the prototype does not serve as the focus of the category, as a basis for organizing instances within the category, as the (analog) representation of the category, or as having any other special status. In this view, the prototype is derivative of feature occurrence information collected schematically by the subject in the course of forming a category.

It is difficult to think of any single experiment which allows a direct comparison of a feature contingency approach to category formation and the prototype approach proposed by Rosch and others. The following experiment, however, sheds some light on the difference between these two formulations. The experiment hinges on what seems to be one of the assumptions of most prototype models, namely that the most typical instance contains all defining characteristics of the concept in question. In other words, the prototype is a conjunction of defining features. There is no room in prototype theory for what might be equivalent to a non-occurrence marker, that is, a tag in memory to indicate that, under some circumstances, feature x_m cannot occur in an exemplar. Suppose the category to be learned is characterized by a set of features, some of which show a low or non-existent conjoint frequency. In other words, features x and y are among the several which show a high probability, contingent relationship with the category. It is the case, however, that if the positive instance has feature x, it is unlikely to have feature y. Alternatively, if it has feature y it is unlikely to have feature x.

Conjoint probabilities are as straightforward to manipulate as are the probabilities of individual features. When the conjoint probability of two defining features is 0, their relationship can be described as an exclusive disjunction: "Members of the category must have (or are likely to have) feature _x_ or feature _y_ but not both." The probability of conjoint occurrence, of course, can be manipulated from 0 through 1.0, the latter extreme being the inclusive disjunctive: "Members of the category must have feature _x_ or feature _y_ or both." The question at issue is, what will the subject designate as the prototypical stimulus and how will he/she judge degree of category membership as conjoint probabilities change from 0 through intermediate values to 1.0. If our reading of prototype theory is correct, even in the case of 0 conjoint frequency (an exclusive disjunctive rule), the subject should choose, as the "best instance," the stimulus which contains both defining features, all other things being equal, and should rank other stimuli for category membership accordingly. In contrast, feature/schema theory predicts no clear prototype or evaluation of instances except that instances with feature _x_ or feature _y_ are equally representative of the category.

Method

Subjects

One hundred and twenty introductory psychology students participated in the experiment in partial fulfillment of a course requirement. The data of three additional subjects were discarded because of experimenter error of apparatus malfunction during a session.

Design

The design was a 5 x 2 x 2 factorial, including five percentage levels for a TT instance positive (0, 25, 50, 75, 100), two stimulus populations (one consisting of 72 unique geometric designs and the other of only 57 from a total of 81 possible designs), and two different problems, each with a different pair of relevant features (one and red versus large and triangle).

All variables were manipulated between subjects, yielding 20 conditions. Six subjects were randomly assigned to each of these cells.

Procedure

Subjects were run individually during a 60 min session. They were told that a series of geometric designs would be presented on the screen in front of them. Each design was either a positive or negative instance of a concept, as indicated by a signal light on a feedback-response panel. The four dimensions of the stimuli and their three values were described: color (red, yellow, black), size (small, medium, large), number (one, two, three), and shape (square, hexagon, triangle). The two relevant dimensions were named for subjects and the general rule for classifying stimuli was described in truth table terms. In effect, subjects were told that one value from each relevant dimension determined stimulus to category assignments, that TF's and FT's (stimuli with only one relevant feature) were always positive, that FF's (stimuli with neither relevant feature) were always negative, and that TT's (stimuli with both relevant features) might be assigned to either category. Subjects were instructed that even though the feedback for TT instances might not appear to be consistent, they should still try to use this information to decide whether or not TT instances were examples of the concept. The subjects, of course, had to determine from feedback which values on the two relevant dimensions defined the concept.

Two practice runs were given, using the alternative concept problem as an example (e.g., if _one_ _red_ was the actual concept, then _large_ _triangle_ was used for the practice trials). During the first practice run, subjects were asked to categorize positive and negative instances, knowing both the relevant features and the rule; no TT's were shown. For the second run, the experimenter explained possible strategies to find the relevant features. These strategies were mainly to use a process of elimination and to look for common features in positive instances.

The rest of the experiment consisted of an acquisition series of six study-test cycles, followed by four post-acquisition tests. The study sequences consisted of eight slides shown for 6 sec each with feedback as to category, positive or negative. No response was made to these slides; the subjects were to use the feedback to try to identify the two relevant features. Two examples from each of the four truth-table classes appeared in each study sequence. The test sequence required the subject to classify four slides, one of each truth table category, as either positive or negative using designated buttons on the feedback-response panel. No feedback was given for test responses.

The following post-acquisition tests were given, in the same order to all subjects.

Typicality rank ordering. Four stimulus cards, one example from each truth table class, were laid before the subject. He/she was asked to rank the cards from the best example of the concept to the worst. Six sets of four cards were rank ordered. Randomized within the six sets were two novel stimuli (not seen during acquisition) from each truth table class.

Speeded classification. The subject was asked to classify 20 slides as examples or nonexamples of the concept learned as rapidly as possible. There were five stimuli, one of them being novel, for each truth table class. A ready signal was given before the first slide; all slides remained on the screen until the subject responded; the inter-slide interval was approximately 1 sec.

Frequency estimates. Subjects were told that 50 positive instances of this concept were presented of during acquisition. They were asked to estimate how many of the 50 (0-50) contained each of 15 named values or value combinations. The six individual values and the nine value pairs from the two relevant dimensions were presented in random order for estimation.

Pairwise comparisons. Subjects were given 24 pairs of stimuli, and were asked to choose the better example of the concept in each pair. The 24 different comparisons represented six truth table combinations, each occurring four times. The six combinations were the product of each truth table class paired with every other class, e.g., TT vs TF.

Apparatus

A Kodak Carousel slide projector was interfaced with solid state circuitry designed to control sequencing, timing, and feedback. Slide sequences and feedback were read from a punched tape. A printout timer recorded all responses and response times.

Results

Acquisition

For the results of post-test to be meaningful, it is important that all subjects, regardless of experimental condition, reach a comparable and stable level of response, at least to those stimuli for which feedback is systematic. In our analysis, we examined responses to TT stimuli separately from the remaining stimulus classes, TF, FT, and FF. Feedback with respect to the proper response for each of the latter three classes was regular for all experimental conditions during study trials. Feedback on TT instances differed across conditions, being stable for Condition 0 (TT instances always negative) and Condition 100 (TT instances always positive), but variable for Conditions

25, 50, and 75. Unless otherwise indicated, all effects described below are reliable at $p < .01$.

Percent correct responses over the six tests during acquisition were analyzed for TF, FT, and FF classes only. There were three significant main effects, Ordinal position of test, F (5, 700) = 21.76, Truth table class, F (2, 1700) = 5.04, and Condition, F (4, 100) = 3.74. The ordinal position effect merely confirms the existence of an acquisition function. On the average, percent correct responses was 47% on the first test and 100% on Test 6. The difference among truth table classes is attributable primarily to performance on FF instances being significantly better than performance on TF and FT instances during early tests. The Condition effect demonstrates that acquisition took place more rapidly when feedback was consistent (Condition 0 and 100) than when feedback was inconsistent (Conditions 25, 50, and 75) for TT instances. Further, Condition 100, in which TTs were called positive 100% of the time, led to more rapid acquisition than Condition 0, in which TTs were called negative 100% of the time.

There were two important interactions, the pertinent data being shown in Tables 1 and 2. First, Class of stimulus instance interacted with Ordinal

Insert Tables 1 and 2

position of acquisition test, F(10,1700) = 3.80. The interaction results from the fact that percent correct response on Test 1 to FF instances was 61% but only 40% for TF and FT instances. Subjects achieved the same level of performance, 100% correct responses, on Test 6. The interaction, then, evidences the convergence of performance functions for truth table classes over trials. Condition interacted with Ordinal position, F(20,1700) = 3.07. This interaction is due essentially to the fact that performance improved more rapidly over tests for Condition 0 and Condition 100, compared to the intermediate conditions. For all groups, percent correct response was approximately 47% on Test 1 and 100% on Test 6. The Condition 0 and 100 diverged and achieved higher performance levels than Conditions 25, 50, 75 on intermediate tests.

An analysis of percent correct response on TF, FT, and FF classes during the last two tests only revealed no significant effects. This outcome suggests that subjects in all conditions and over all stimulus classes achieved the same acquisition level and the same understanding of the concept, at least as it applies to stimuli for which feedback was consistent.

Over the entire series of acquisition tests, TTs were called positive instances .35, .59, .70, .80, and .96 of test trials by subjects in Conditions 0, 25, 50, 75, and 100, respectively. These differences are highly reliable, $\underline{F}(4,100) = 5.44$. There is a strong initial tendency to call TTs positive by subjects in all conditions, the proportion averaging .81 on the first test administered. Subjects learned, however, to adjust this response in light of feedback provided over study trials. Thus, on the last two tests, the proportion of TTs called positive was .01, .31, .59, .79, and 1.0 by subjects in Condition 0, 25, 50, 75, and 100, respectively. These differences were also statistically reliable, $\underline{F}(4,100) = 14.80$. In an analysis of these data, Ordinal position of tests, $\underline{F}(5,500) = 3.17$ and the interaction of Conditions by Ordinal position, $\underline{F}(20,500) = 5.44$, were statistically significant. These performance functions are reminiscent of probability matching functions observed in a variety of simpler learning tasks (Estes, 1976).

Post tests

Different stimulus populations were sampled during acquisition to provide differing degrees of access to old or previously seen vs. new instances on the post-tests. Throughout the analyses of post test data, no differences were detectable between these populations or between new and old stimuli, possibly because so few novel stimuli were available from either population. But, because there were no differences, all data and analyses reported will ignore the differences between old and new stimuli and between stimulus populations.

Typicality rank ordering. For each subject, we computed the average rank of all TT, TF, FT and FF instances presented in the six sets of four cards which subjects were asked to rank order. In an analysis of variance on these data, the difference among Conditions contributed practically no variance, as one would expect given that the sum of rank was a constant across conditions. The difference among classes of stimuli was highly significant with TT stimuli, on the average, assuming the top rank, FF stimuli falling at the lowest rank and TF and FT stimuli ranking in between, F(3,300) = 966.26. Large as it is, the effect is rendered relatively uninteresting by the existence of a Stimulus class by Condition interaction, F(12,300) = 3.18, shown in Table 3. FF stimuli obtained a relatively constant and low rank across conditions. The mean rank of TT instances increased and, compensatingly, the rank of TF and FT instances, which do not differ, decreased across conditions. For Condition 100, subjects were almost entirely consistent in their ranks, placing TT at the top, FF at the bottom, and TF and FT at Ranks 2 and 3. In Condition 0, TF and FT obtained, on the average, tied ranks with TT instances next and FF instances ranked lowest. Ranks changed systematically as the percent of TTs in the positive category increased. Thus, even with relatively modest feedback to the effect that

Insert Table 3

TT instances are sometimes positive, subjects, on the average, ranked TT instances as the best example of the concept that had been learned.

Speeded classification. Performance on this task was relatively unenlightening. A correct classification of TT instances would place them in the negative category in Condition 0 and in the positive category in Condition 100. The proper classification of TT instances for intermediate conditions is indeterminant. As in the final two tests administered during acquisition, however, subjects tended to divide TT instances between the positive and negative category according to the percentages of feedback administered during acquisition. Error rates can be determined only for Condition 0, .1, and Condition 100, .03 for TT instances. The proportion of TT instances called positive was .31, .57, and .81, for Conditions 25, 50, and 75, respectively. Reaction time to TT stimuli on correct response trials in Condition 0 (TT negative) and in Condition 100 (TT positive) was 3.86 and 3.21 seconds, respectively. Reaction times taken over all trials showed no significant trend attributable to conditions. Reaction time in Condition 100 (3.41 sec) was, however, significantly faster than the pooled reaction times for the remaining conditions (3.75 sec), $\underline{F}(1,100) = 3.52$.

Frequency estimates. Subjects were required to estimate on a scale from 0 to 50 how often each of fifteen features or feature combinations had occurred in positive instances during acquisition. All individual features and combinations on the two relevant dimensions were presented for estimation. The mean obtained values over five conditions are presented in Table 4. In an analysis of variance of these data, there were two effects of significance, Features (or Feature combination), $\underline{F}(14,1400) = 160.86$, and the interaction of Features and Conditions, F(56,1400) = 3.32. The meaning of these effects is readily apparent from Table 4. Frequency estimates increase systematically over

Insert Table 4

Conditions 0 through 100 for the individual relevant features and for the combination of the two relevant features, the combination showing a steeper rise because of the lower estimate in Condition 0. Complementary trends are demonstrated for non-relevant individual values on relevant dimensions and for combinations of 1 relevant and 1 irrelevant value. Where neither value of a pair was relevant, estimates on the average were close to 0.

These estimates are essentially what one would expect, given that all subjects mastered the concept during acquisition. It is potentially of some importance to ask how accurate these estimates were and whether or not accuracy correlates in any way with concept acquisition. We constructed an accuracy measure by converting frequency estimates to a probability and then taking the difference between estimated and actual probability of a certain event, given the sequence of acquisition stimuli presented. Of major interest in the analysis of these data are the correlations between accuracy of estimation and number of errors committed on acquisition test trials. Two correlations were statistically reliable, one based on the occurrence of either relevant feature, $r = .390$, $t(118) = 3.17$, and the other on the occurrence of both relevant features, $r = .330$ $t\ (118) = 2.95$, suggesting that the accurate accumulation of frequency information bears some relationship to the ability to form relatively complex concepts.

Pairwise comparisons. We used as primary data the number of times each member of a particular pair was chosen as the better example of the concept learned. Reducing the various stimuli to their truth table classifications, there are six possible comparisons. Those six comparisons were analyzed independently in cannonical order as a within subjects variable. For purposes of more efficient comparison, however, we combined certain pairs in the presentation of data in Table 5. To be specific, because there was no reason to expect any difference between the TT-TF and the TT-FT comparison or between the FF-TF and the FF-FT comparison, these pairs are combined. The values presented in Table 5 reflect the average frequency with which the first member of each pair was selected as more exemplary of the concept. In the analysis of variance, three effects of interest were statistically reliable: Condition, $\underline{F}(4,100) = 3.58$, Pair, $\underline{F}(5,500) = 131.83$, and the interaction of these variables, $\underline{F}(20,500) = 2.77$. These data can be described as follows. Subjects are indifferent to the comparison between TF and FT stimuli, choosing one or another with roughly equal frequency regardless of experimental conditions. Also regardless of condition, subjects picked with high frequency either the TF or the FT as a better example than FF stimuli. Condition, as one would expect, produces its major effect on comparisons involving TT stimuli. Even in Condition 0, TT stimuli are picked as better examples than FF stimuli although both are equally non-exemplary of the concept. This preference for TT stimuli increases as a percent of TTs placed into the positive category increases across conditions. In the comparison of TT with TF or FT stimuli, there is an important crossover. In Condition 0, TF or FT stimuli are chosen more frequently as better instances than TT, although the choice of TT as a better instance has some significant value even in Condition 0 where TT instances are always negative. The point of indifference between TT and TF or FT instances lies between Conditions 50 and 75, despite the fact that TF and

----

Insert Table 5

FT instances are always positive and TT instances are positive only 50-75% of the time. In Condition 100, TT instances are almost always chosen in preference to TF and FT instances even though all three instance classes have appeared with the same regularity in the positive category.

Discussion

A concept based on some well-defined rule, even a rule that integrates only a pair of relevant stimulus features, yields phenomena not unlike those observed by Rosch (1978) for so-called natural concepts. Subjects in Conditions 0 and 100 of the present experiment learned easily and to a high criteria to place stimuli into positive and negative categories based on a logical rule that integrates two stimulus features. Despite their training and their knowledge of the rule, subjects in Condition 0 nevertheless exhibited a tendency to respond to a stimulus with both features as a better example than a stimulus with neither feature. That tendency was even more obvious in other conditions where at least some TT instances were assigned to the positive category. The potency of this conjunction of features is clearly revealed in various post-test performances even when only a small percentage of stimuli with both critical features is called positive during training.

Thus, we conclude that, even when the subject learns a well defined rule based on primitive stimuli, there are underlying processes which generate performance superficially reflecting a dimension of instance goodness. The processes, we argue, derive from the formation of a concept-specific schema during training. The subject enters the problem with a generalized schema containing slots which act in the task as requests for information. Some of the slots are identified by known stimulus dimensions, established during careful and detailed instructions given at the outset. One kind of information, but by no means the only kind, that can be requested and recorded in these slots corresponds to category contingencies or frequency of feature occurrence within categories. As the subject records this information, dimensional distinctions begin to emerge and to determine feature relevance. Because individual stimuli differ in the combinations of features they contain and because positive stimuli need not, in general, include all relevant features, a quality of instance goodness can be defined directly. Such an underlying set of principles, as outlined earlier, leads directly to the phenomena of "best instance" and typicality which appear ubiquitously in studies of natural concepts. Schema theory implies that it is not concept fuzziness that generates these observations. Rather, it is the application of a clear and parametrically well-defined concept to instances whose feature structure varies so as to produce these effects.

Contrary to Rosch (1978), we argue that simplicity in concepts or concept tasks does not produce an unnatural phenomenon. Rather, the phenomena are essentially the same in natural and logical concepts. Furthermore, we claim that the notion of typicality, rather than being fundamental to concepts, is derivative of more basic processes which can best be examined in the laboratory.

Footnote

1. This research was conducted within the Institute for the Study of Intellectual Behavior. The work was supported by Research Grant BNS 76-81416 from the National Science Foundation. The author wishes to thank Michelle Collins, Linda Liebelt, and Peder Johnson for various contributions to the project, ranging from construction of stimulus materials through extremely useful discussions about the theoretical ideas expressed. Of course, any errors of execution or logic are solely the responsibility of the author.

References

Estes, W.K. The cognitive side of probability learning. *Psychological Review*, 1976, *83*, 37-64.

Garner, W.R. Aspects of a stimulus: Features, dimensions and configurations. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. New York: Wiley & Sons, 1978.

Rosch, E. Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. New York: Wiley & Sons, 1978.

Rosch, E. & Mervis, C.G. Family resemblances: Studies on the internal structure of categories. *Cognitive Psychology*, 1975, *7*, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382-439.

Table 1

Probability of a Correct Response Averaged across Conditions during Acquisition

Truth Table	Ordinal Position of Test					
Class	1	2	3	4	5	6
TF	.405	.533	.732	.866	.975	1.000
FT	.393	.518	.755	.838	.985	1.000
FF	.605	.714	.819	.895	1.000	1.000

Table 2

Probability of a Correct Response on FT, TF, and FF Instances

During Acquisition Ordinal Position of Test

Condition	1	2	3	4	5	6
0	.485	.591	.787	.869	.995	1.000
25	.477	.573	.726	.822	.983	1.000
50	.459	.567	.728	.808	.977	1.000
75	.468	.500	.752	.822	.987	1.000
100	.475	.683	.808	.990	1.000	1.000

Table 3

Typicality Rank Order of Stimulus Instances across Conditions

Stimulus Class	Condition 0	25	50	75	100
TT	2.89	1.58	1.32	1.11	1.00
TF	1.81	2.12	2.21	2.58	2.52
FT	1.72	2.25	2.31	2.32	2.51
FF	3.83	3.99	4.00	3.98	3.97

Table 4

Mean Estimates of Frequency within Positive Instances

	Single features		Two feature combinations		
Conditions	Relevant	Irrelevant	Both relevant	One relevant	Both irrelevant
0	30.2	19.1	5.7	39.0	0.5
25	37.2	20.6	20.5	45.2	0.0
50	43.7	20.0	30.9	48.5	1.3
75	47.7	18.8	41.6	49.2	0.0
100	49.5	19.5	50.0	48.8	0.8

Table 5

Average Number of Times the First Member of Each Pair Was Chosen as the Better Example of the Concept

Pair	Condition				
	0	25	50	75	100
TT-FT(FT)	1.091	1.625	2.137	2.833	3.542
TT-FF	2.954	3.500	3.875	4.000	4.000
TF-FT	2.037	2.208	2.054	1.833	1.922
TF(FT)-FF	3.708	3.792	3.955	3.833	3.542