Are rating-based procedures reliable for derivation of typicality judgments from children?

KATERINA MARIDAKI-KASSOTAKI
University of Crete, Rethymnon, Crete, Greece

The aim of the present study was to examine whether a rating-based procedure that has already been used by other investigators can be used for derivation of typicality ratings from children. The evidence reported in the study establishes that such a procedure cannot be reliably used for this purpose. The results show that children rated category items in terms of personal preferences rather than as a function of how representative they considered the items to be of their superordinate category. On the basis of these findings, an alternative method based on the family resemblance scores of the category members was proposed in order to derive typicality ratings from young children. This family resemblance method of obtaining typicality judgments may be useful to investigators interested in assessing how children process categorical information.

The method widely used by investigators (e.g., Rosch, 1973, 1975) to derive typicality ratings requires subjects to judge the degree of typicality of a particular category member on the basis of a rating scale. Normally, subjects are asked to rate items according to a 7- or 6-point scale, ranging from 1 for *very good* category members, through 2, 3, 4, 5 for *quite good, fairly good, good*, and *poor*, respectively, to 6 or 7 for *very poor* category members.

It is not possible to employ such a rating task with very young children as subjects; the linguistic and cognitive demands of the task are too difficult. Previous investigators have, however, used 3-point rating scales to derive typicality ratings from children in a playful context (see Bjorklund, Thompson, & Ornstein, 1983). In this study, children were asked to rate category items in order to help a person from another planet understand the meaning of the categories. The task required the children to rate the items on a 3-point scale in a playful context. Three schematic faces, each having a different expression, were presented to them. The children were told to rate the items while pointing to one of the schematic faces. The "smiling" face corresponded to "very good" category members. These members would be very helpful to the spaceman for understanding the category. The "straight" face corresponded to "okay" category members. These members would help the spaceman somewhat to understand the category. The "frowning" face corresponded to "poor" exemplars, items that would not help the spaceman understand the category very well. The children were instructed to rate the items in terms of how good an example each item was of its category and not in terms of how much they liked the item.

The findings obtained by Bjorklund et al. (1983) provided strong support for the position that the boundaries of categories expand with age. They also confirmed the observations of other researchers that adult-defined typical category members become part of children's categories before atypical members do. A problem, however, related to the procedure adopted in this study arises from the way in which children understand the words *good* and *poor*. The question "How good and/or poor is a [name of an exemplar] of [name of the corresponding category]?" can be interpreted by children in a way different from that assumed by Bjorklund and collaborators. Children may choose, for example, to answer this question in terms of personal preferences rather than the degree to which an item is representative of its category. Thus, for example, a child who dislikes dogs may judge them to be "bad" animals. It is, however, questionable whether the procedure of asking children to judge exemplars as "good" or "poor" is likely to yield interpretable results.

In a second experiment, Bjorklund et al. (1983) tested such a possibility with a group of age mates of the subjects involved in the original experiment. This time, the subjects were asked to rate the items on a 3-point scale according to "likes" and "dislikes." The typicality ratings produced on the basis of children's "likes" and "dislikes" were correlated with the typicality ratings derived from children's judgments of "good" and "poor" on a 3-point rating scale. The low value (.46) obtained from these analyses led the researchers to suggest that young children's judgments of typicality produced on the basis of a 3-point scale were not influenced by their personal experiences.

There is, however, another way to view the results obtained in Bjorklund et al.'s (1983) second experiment. If, as noted earlier, the children actually rated the items by using an affective criterion in Bjorklund et al.'s first experiment, the typicality ratings that they produced should not be highly correlated with typicality ratings derived in terms of "likes" and "dislikes" from a group of age

mates. This is because the use of an affective criterion in judgments of typicality is not stable across subjects and may explain the differences in typicality ratings for the same items between age mates. Therefore, it is also likely that the low values obtained from the correlations in the second experiment of Bjorklund et al. reflect differences in "likes" and "dislikes" between age mates (see also Maridaki-Kassotaki, 1993).

The present study was designed with these methodological considerations in mind in order to explore whether the procedure used by Bjorklund et al. (1983) could be reliably used for derivation of typicality ratings from children. Three experiments were conducted. The aim of Experiment 1 was to generate typicality ratings from young children for items belonging to each of four categories with the use of the procedure employed by Bjorklund et al. for the same purpose. An additional goal of Experiment 1 was to derive ratings for the same category items on the basis of the subjects' like/dislike judgments. The prediction made was that typicality ratings obtained with the use of Bjorklund et al.'s (1983) procedure would correlate significantly with ratings derived in terms of "likes" and "dislikes." To anticipate the data, the children's ratings generated with Bjorklund et al.'s procedure were highly related to their ratings for liked/disliked objects.

In Experiment 2, typicality ratings were obtained from young children by means of an alternative method. This method was based on the procedure employed by Rosch and colleagues (Rosch & Mervis, 1975) for obtaining the family resemblance typicality of a category. Experiment 3 was designed to examine whether the family resemblance method is a reliable way of eliciting typicality ratings from children. The aim of the experiment was to test the correspondence between adults' family resemblance and typicality ratings on the same materials. An additional goal of Experiment 3 was to compare children's ratings derived according to Bjorklund et al.'s (1983) technique to adults' family resemblance and typicality ratings for the same category items.

EXPERIMENT 1

Method

Subjects. A total of 96 kindergarten children participated in the present study. Mean ages were 5 years 1 month (4,3–6,0), $SD = 0.61$). The children were selected from three public kindergartens in Athens, Greece, and were from predominantly middle-class homes.

Materials and Design. The materials were black-and-white drawings depicting 12 items from each of four categories: animals, fruits, furniture, and clothes. Each item was mounted on 10 × 15 cm cardboard. The items were generated in a separate study designed to provide appropriate ranges of category exemplars for young children. In this study, 40 children aged 4–5 years stated as many instances of each of four categories as they could in 10 min. In the present study, the list for each of the four categories comprised 12 items selected on the basis of their production frequency. These items are listed, in their production frequency order, in the Appendix. In addition, three noncategory items were included in the item list of each category as foils.

There were also three pictures, each of which depicted a schematic face with a particular expression. Each picture was mounted on 25 × 25 cm cardboard. The first picture showed a "smiling face"; the second, a "straight face"; and the third, a "frowning face." All the children were tested with all 48 stimuli. The presentation order of the four categories was randomized for each child. The items presented to each child were randomly selected within each category list. Subjects were asked to rate each item on a 3-point scale in terms of "how typical" it was of its category. They were also asked to make like/dislike judgments on the basis of a 3-point liking scale for all category items. The typicality and like/dislike ratings were counterbalanced across subjects.

In addition, each child was presented a practice category that had four items varying in typicality. The practice category, *flowers*, comprised the items *roses, carnations* (good exemplars), *daffodils, daisies* (moderately typical exemplars), and *pansies, fuschia* (atypical exemplars). The typicality of these items was defined by a group of adults on the basis of a 6-point rating scale. The practice category was introduced to ensure that the subjects understood the task demands.

Procedure. Each child was tested individually in a room in which the experimenter and the child were alone. The experimenter sat facing the child with a table between them, on which were a toy spaceship and two toy dolls showing two extraterrestrial persons and also the pictures of three schematic faces. The first picture depicted a "smiling face," the second a "straight face" and the third, a "frowning face." The children were read the items of each category and were asked to mention whether these items were members of the category named by the experimenter. Then they were asked to rate the practice items on a 3-point scale with assistance from the experimenter. The following instructions were given to each child:

Hello, ... This is Nikos and Niki. They are brother and sister. Look at their spaceship [the experimenter points to the toy spaceship]. It has just arrived from the planet Mars. Nikos and Niki do not know anything about our world. They both want to learn so when they go back they will tell their friends. They also want us to help them learn. Look at these pictures [the experimenter points to the schematic faces]. They will help us to tell Nikos and Niki what they want to learn about.

[The experimenter points to the drawings of roses and carnations.]

Roses and carnations are "very good" examples of flowers. Okay? To tell Nikos and Niki that roses and carnations are "very good" examples of flowers we will point to this picture [the experimenter shows the picture with the "smiling" face on it]. Now listen carefully. "Very good" examples of flowers are ones that might come to mind immediately when thinking of flowers, and ones that would be very helpful to Nikos and Niki to learn about flowers.

[The experimenter points to the drawings of daffodils and daisies.]

Daffodils and daisies are "okay" examples of flowers. Okay? To tell Nikos and Niki that daffodils and daisies are "okay" examples of flowers we will point to this picture [the experimenter shows the picture with the "straight" face on it]. Now listen carefully. "Okay" examples are those that could help the spaceman somewhat to learn about flowers but not as well as would the "very good" items (such as "roses" and "carnations").

[The experimenter points to the drawings of pansies and fuchsia.]

Pansies and fuchsia are "poor" examples of flowers. Okay? To tell Nikos and Niki that pansies and fuchsia are "poor" examples of flowers we will point to this picture [the experimenter shows the picture with the "frowning" face on it]. Now listen carefully. "Poor" examples are those that would not help Nikos and Niki learn about flowers very well.

After the experimenter made sure that each child had understood the task demands, the whole procedure was repeated with the four test categories. Before each child was introduced to the test categories, the experimenter explained to the child that he/she was to say how good an example each item was of its category and not to

tell the experimenter how much he/she liked the items. To make sure that children understood what they were instructed to do, "souvlaki" was used as an example of a Greek food that they might like very much but that was a fairly poor example of "food."

A day after this testing had taken place, the subjects were asked to make like/dislike judgments for all category items. The procedure was the same as the one followed for the derivation of typicality judgments on the basis of Bjorklund et al.'s (1983) technique. The only exception was that the items were presented in different order and the subjects were instructed to tell the two spacemen how much they liked the category exemplars. Thus, children were asked to state whether they (1) "liked very much," (2) "simply liked," or (3) "did not like at all" each category item.

The procedure used in this study was basically the same as that employed by Bjorklund et al. (1983), with slight modifications. The modifications concerned the way in which subjects were asked to make typicality judgments. The subjects in Bjorklund et al.'s study were read the lists of items, whereas in the present study, the subjects were shown cards with black-and-white drawings of the items on them. The results obtained in a pilot study guided the selection of drawings. In the pilot study, it was observed that children were easily distracted and were reluctant to participate when they were read the lists of items as in Bjorklund et al.'s procedure. For this reason, a study was designed to examine whether an alternative way of testing the subjects would be appropriate for the purpose of the present study. It was assumed that the use of visual materials might motivate children to participate. In this study, a group of children ($N = 18$) who were age mates of the subjects participating in the present study were tested. The subjects were asked to make typicality judgments while (1) listening to the experimenter name the items and (2) being shown cards with drawings of the items on them. Correlations between the ratings obtained with the experimenter reading the item lists and those generated with the use of drawings were significant ($r = .78$ for animals, $r = .82$ for fruit; $r = .78$ for clothes and $r = .86$ for furniture, $p < .001$). On the basis of this evidence, a decision was made to present black-and-white drawings of the items to the children rather than to read the items to them.

Results and Discussion

Each item received a score ranging from 1 (selected as a *poor* category member) to 3 (selected as a *very good* category member). The score of 2 corresponded to *okay* category members. Similarly, a value of 3 corresponded to a response of "I like [name of an exemplar] very much," and the values of 2 and 1 to the responses "I simply like [name of an exemplar]" and "I don't like [name of an exemplar] at all," respectively.

The scores generated for an item on the basis of the 3-point rating scale across subjects were summed, thus yielding the typicality score for this item. The higher the score for an item, the greater its judged typicality. Similarly, the scores generated for an item on the basis of like/dislike judgments across subjects were summed, thus yielding the degree of liking score for this item. The typicality and degree of liking scores of the items from each of the four categories are presented in Table 1.

Pearson correlation coefficients were computed between typicality and degree of liking scores across categories. The correlations between the typicality ratings produced on the basis of Bjorklund et al.'s (1983) procedure and the ratings generated in terms of "likes" and "dislikes" were very high. These correlations were as fol-

Table 1
Category Items and Their Typicality Scores and Degree of Liking Scores: Experiment 1

Category	Typicality	Liking
Animals		
Cat	242	240
Dog	236	245
Cow	211	210
Chicken	180	204
Lion	158	165
Pigeon	155	169
Goat	153	146
Monkey	146	129
Turtle	136	143
Wolf	128	109
Bee	119	110
Snake	110	101
Fruit		
Apple	170	168
Mandarin	164	176
Orange	152	151
Pear	144	143
Strawberries	135	141
Cherries	128	138
Banana	127	131
Apricot	125	125
Pomegranate	116	128
Grapes	115	135
Pineapple	113	108
Fig	111	120
Clothes		
Trousers	162	156
Dress	156	132
Skirt	148	149
Scarf	140	142
Blouse	134	130
Shirt	132	122
Socks	128	125
Anorak	127	119
Cardigan	120	121
Hat	116	109
Coat	101	111
Gloves	96	106
Furniture		
Chair	185	160
Sofa	170	190
Bed	158	175
Desk	150	149
Table	145	158
Stool	139	144
Lamp	135	137
Bookcase	130	140
Wardrobe	124	119
Picture	120	130
Cupboard	116	110
Bedside table	110	126

lows: animals, .97; fruit, .93; clothes, .89; furniture, .85. All were significant ($ps < .001$).

Such results strongly confirm our hypothesis that typicality ratings derived for category items with the use of Bjorklund et al.'s (1983) procedure should be highly correlated with ratings obtained in terms of "likes" and "dislikes." From the evidence available here, it seems clear

that children's typicality ratings reflect their personal preferences rather than the representativeness of each exemplar within its category. With respect to the present findings, therefore, it may be concluded that the procedure employed by Bjorklund and colleagues to derive typicality ratings from children cannot be reliably used for this purpose.

EXPERIMENT 2

The results from Experiment 1 indicate that reliable indications of the typicality structure of categories cannot be obtained by asking young children to judge the goodness of correspondence between an item and its superordinate category. Children 4–5 years of age failed to respond adequately to the demands of a rating task, despite the fact that the rating procedure was simplified and the instructions were given in a playful situation. It was suggested that the way in which children interpreted the questions misled them to wrong answers.

The results reported in Experiment 1 identify the need for an alternative method to derive ratings of category items from young children. The goal of the present study was to derive typicality ratings with the use of a procedure that has been used so far with adults, but not with children, for this purpose (e.g., Rosch & Mervis, 1975). Most specifically, this experiment was designed to provide a portrait of the family resemblance scores for four categories, each consisting of the items used in Experiment 1. The family resemblance scores of the category items were used as a basis for generating the typicality ratings of these items.

Research on the internal structure of categories has suggested that the members of categories are organized by *family resemblances*. On the family resemblance view, not all members of a category possess exactly the same features. Although a category member may share one or no features with one or more members, no one or few features must be shared by all members in the category (Gleitman, Armstrong, & Gleitman, 1983; Hampton, 1979; Rosch & Mervis, 1975; Smith & Medin, 1981). Furthermore, in a number of experiments Rosch and her colleagues have shown that the family resemblance scores calculated for each item by weighting the features that it possesses across all category members are highly correlated with typicality ratings provided by individuals for these items. On the basis of these findings, it was concluded that the more attributes an item shares with other items of a category, the more representative (typical) member of its category it is (Rosch & Mervis, 1975). The distribution, therefore, of the features of the members of a category can be used as a basis for deriving typicality ratings for these items. In Rosch and colleagues' studies, the procedure employed for obtaining the family resemblance scores of the members of a category was straightforward. Subjects were asked to describe items in terms of attributes such as "has legs," "is edible," "is made of wood," and so forth. The advantage of this technique over the rating procedure applied in Experiment 1 is that it may

provide a simple means for deriving typicality ratings with children. Even very young children seem likely to be able to generate features. Because family resemblance scores of category items are good predictors of typicality ratings for these category items (see Rosch & Mervis, 1975), it should be possible to establish typicality ratings for children with the use of the family resemblance structure of the categories.

There are two reasons, therefore, why family resemblance scores were used for selecting category ratings: first, family resemblance scores are good predictors of typicality ratings, and second, young children are able to respond to the task employed for generating family resemblance scores, because it is simple and straightforward. It was assumed that if typicality ratings generated either with the procedure employed by Bjorklund et al. (1983) or based on the degree of liking did not correlate highly with family resemblance ratings, it would be necessary for us to use the procedure that leads to the derivation of the family resemblance scores to generate typicality ratings from young children.

In Experiment 2, all subjects were asked to generate attributes for the members of four categories for the benefit of two people from another planet who had no knowledge about the particular items. The items of each category were ranked on the basis of their family resemblance scores. The higher the family resemblance scores of the items, the more typical these items were of their category.

Method

Subjects. The children who participated in Experiment 1 served as subjects in Experiment 2.

Materials and Design. The materials were the same as those used in Experiment 1. Each item was given a featural definition by 8 subjects. Each subject provided featural definitions for four items, one from each category. Items were randomly selected within their categories, so that each subject who received an item received it with different items from the other three categories and received the items representing each category in a different order. All subjects were tested 2 weeks after Experiment 1 had been carried out.

Procedure. Each child was tested individually in a room in which the experimenter and the child were alone. The experimenter sat facing the child with a table between them, on which were a toy spaceship and two toy dolls showing two extraterrestrial persons. The experimenter then gave the following instructions to each child:

Hello ... This is Nikos and Niki. They are brother and sister. Look at their spaceship [points to the toy spaceship]. It has just arrived from the planet Mars. Nikos and Niki do not know anything about the planet Earth. They both want to learn so when they go back they will tell their friends about our world.

[At this point the experimenter takes Nikos to her ear and pretends that he is whispering to her. Then she turns to the child and tells him/her:]

Nikos has just told me that he very much wants to learn about our world. Will you help them?

[Nikos whispers again. The experimenter turns to the child and says:]

Nikos has just told me that he does not know what a dog is and he wants you to tell him everything you know about dogs.

Will you please try hard and tell him everything you know about dogs?

[The child gives an answer. Nikos whispers again. The experimenter turns to the child and says:]

**Table 2
Attributes Applied to the Fruit Item *Pear*
by a 5-Year-Old Child: Experiment 2**

Item	Attributes
Pear	It is a – fruit – (2)
	It is – green – (1)
	It has – skin – (14)
	It has – vitamins – (28)
	It has – a stem – (3)
	We – peel – it (13)
	We – eat – it (4)

Note—See text for explanations.

Nikos has just told me that he really enjoys listening to you but he still does not understand what a dog is and he wants you to tell him what a dog (chair, etc.) is.

Will you please try and tell him what a dog (chair, etc.) is?

[The child gives an answer. Niki whispers now. The experimenter turns to the child and say ·]

Niki has just told me that she could not understand what dogs (chairs, etc.) have and she wants you to tell her what dogs (chairs, etc.) have.

Will you please try and tell her what a dog (chair, etc.) has?

In order to encourage the subjects to generate as many properties as possible, the subjects were prompted to do so every 5 min. The questions comprising the instructions were counterbalanced across subjects. The children's answers were recorded on an audio cassette. The procedure used in the present study was very similar to the one developed by Rosch and Mervis (1975). The only exception was that the instructions were given in a playful context. Thus, the children were motivated to participate and provide appropriate answers.

Results and Discussion

The results are presented in two sections. The first section outlines how the attributes applied to each category member were produced. The second section reports how the family resemblance scores for each category member were constructed.

Scoring category attributes. Following Rosch and Mervis (1975), every attribute applied by all 8 subjects to each category item was classified as a separate unit and was assigned a unique integer that served as its nominal code in later analyses. An attribute was defined as the smallest knowledge unit that could not be decomposed into other attributes. The attributes that applied to all category items composed the feature list for their category.

As an illustration, Table 2 shows the attributes given to the fruit item *pear* by a 5-year-old child. In this table, the words set off with dashes indicate the attributes applied to the item *pear*, and the number in parentheses attached to each attribute shows the integer that this particular attribute obtained in its feature list.

Two judges received the feature lists and indicated cases in which the same attribute appeared in different phrasings. Those attributes were combined into a single attribute. For the category of animals, for instance, the properties *pretty/beautiful/nice* were combined into the property *beautiful*. Other phrases such as *they keep us warm* and *they keep us cool*, found in the category of clothes, were considered as a single attribute. Further cases in which subjects listed opposite but not contradicting attributes

**Table 3
Fruit Items and their Family Resemblance Scores: Experiment 2**

Items	Attributes						Family Resemblance
Pear	$A_1(7)$ $A_3(7)$ $A_6(5)$ $A_{13}(6)$ $A_{28}(5)$						75
	$A_2(12)$ $A_4(9)$ $A_7(9)$ $A_{14}(7)$ $A_{30}(7)$ $A_{41}(1)$						
Strawberries	$A_1(7)$ $A_3(7)$ $A_7(9)$ $A_{26}(6)$ $A_{49}(1)$						60
	$A_2(12)$ $A_4(9)$ $A_{13}(6)$ $A_{46}(1)$ $A_{50}(1)$ $A_{42}(1)$						
Apple	$A_1(7)$ $A_3(7)$ $A_5(5)$ $A_{14}(7)$ $A_{30}(7)$						70
	$A_{51}(1)$ $A_2(12)$ $A_4(9)$ $A_7(9)$ $A_{17}(2)$						
	$A_{52}(1)$ $A_{53}(1)$ $A_{54}(1)$ $A_{60}(1)$						
Mandarin	$A_2(12)$ $A_5(5)$ $A_7(9)$ $A_{26}(6)$ $A_4(9)$						56
	$A_6(5)$ $A_{13}(6)$ $A_{35}(4)$						
Fig	$A_1(7)$ $A_5(5)$ $A_{14}(7)$ $A_{30}(7)$						51
	$A_2(12)$ $A_7(9)$ $A_{17}(2)$ $A_{38}(1)$ $A_{58}(1)$						
Grapes	$A_4(9)$ $A_6(5)$ $A_{14}(7)$ $A_{30}(7)$ $A_{57}(1)$						72
	$A_3(8)$ $A_2(12)$ $A_5(5)$ $A_{13}(6)$ $A_{28}(5)$						
	$A_{35}(4)$ $A_{59}(1)$ $A_{61}(1)$ $A_{63}(1)$						
Cherries	$A_4(9)$ $A_6(5)$ $A_{14}(7)$ $A_{30}(7)$ $A_3(7)$						68
	$A_5(5)$ $A_{13}(6)$ $A_{28}(5)$ $A_{35}(4)$ $A_2(12)$ $A_{66}(1)$						
Orange	$A_1(7)$ $A_6(5)$ $A_7(9)$ $A_{30}(7)$						55
	$A_4(9)$ $A_{13}(6)$ $A_2(12)$						
Pineapple	$A_1(7)$ $A_7(9)$ $A_{28}(5)$ $A_{35}(4)$						46
	$A_2(12)$ $A_{26}(6)$ $A_{34}(1)$ $A_{32}(1)$ $A_{44}(1)$						
Apricot	$A_4(9)$ $A_{26}(6)$ $A_3(7)$ $A_{14}(7)$						49
	$A_{30}(7)$ $A_2(12)$ $A_{71}(1)$						
Pomegranate	$A_3(7)$ $A_7(9)$ $A_1(7)$ $A_4(9)$ $A_{39}(1)$						58
	$A_{26}(6)$ $A_{28}(5)$ $A_{43}(1)$ $A_2(12)$ $A_{64}(1)$						
Banana	$A_7(9)$ $A_3(7)$ $A_{14}(7)$ $A_{56}(1)$						43
	$A_2(12)$ $A_{26}(6)$ $A_{70}(1)$						

Note—The number in parentheses indicates how often a particular attribute (A) occurs in the 12 fruit items. See text for further explanation.

for a particular item were treated as the same attribute. For example, one subject listed *not wild* as an animal attribute, and another subject, *tame*. Those attributes were combined into a single attribute. Similar cases were treated accordingly.

Deriving the family resemblance scores. Every attribute listed for an item was weighted according to the total number of items in the category that possessed that attribute. Each attribute, therefore, received a score ranging from 1 to 12, which represented the number of items in the category that possessed that attribute. Table 3 provides an illustration of the family resemblance analysis based on the protocols generated by the subjects for the category of fruit.

In Table 3, the number in parentheses indicates how often a particular attribute occurs in the 12 fruit items. For example, $A_2(12)$ shows that the attribute 2, which refers to "it is a fruit," is shared by 12 fruit items. The measure of the family resemblance of an item was the sum of the weighted scores of each of the attributes that had been applied to that particular item.

The items of each category were ranked on the basis of their family resemblance scores. Following Rosch and Mervis (1975), the higher the family resemblance score of a category member, the more it was considered a good and representative member of its category. Table 4 presents the items from each category, listed, in their typicality order, on the basis of their family resemblance scores.

To test the hypothesis that there is no correspondence between typicality ratings produced on the basis of family resemblance scores and typicality ratings generated in terms of the Bjorklund et al. (1983) procedure and also ratings generated on the basis of likes and dislikes, correlations were performed between family resemblance scores, typicality à la Bjorklund scores, and degree of liking scores. The results are presented in Table 5.

Table 5 clearly shows that the correlations between the three measures were very low across categories. They all failed to reach significance ($p < 1$). The family resemblance scores, therefore, did not correlate significantly with either the typicality scores obtained with the use of Bjorklund et al.'s (1983) procedure or the ratings generated in terms of "likes" and "dislikes." These results corroborate the hypothesis that typicality ratings produced on the basis of the family resemblance scores are not good predictors of typicality ratings obtained with Bjorklund et al.'s procedure or of affective ratings based on like/dislike judgments. The present findings, therefore, suggest that the family resemblance scores generated in terms of the featural distribution of the category members as proposed by Rosch and Mervis (1975) can be used as a basis to derive typicality ratings from young children.

EXPERIMENT 3

To ensure that family resemblance judgments are good measures of typicality, Experiment 3 was conducted, in which the reliability of the family resemblance method for deriving typicality ratings was examined by compar-

Table 4
Category Items Listed in Order of Their
Family Resemblance Scores: Experiment 2

Category	Family Resemblance
Animals	
Lion	99
Dog	95
Wolf	90
Cat	87
Goat	84
Cow	81
Monkey	77
Chicken	73
Turtle	65
Bee	59
Snake	50
Pigeon	45
Fruit	
Pear	75
Grapes	72
Apple	70
Cherries	68
Strawberries	60
Pomegranate	58
Mandarin	56
Orange	55
Fig	51
Apricot	49
Pineapple	46
Banana	43
Clothes	
Anorak	78
Blouse	75
Skirt	74
Trousers	69
Dress	61
Coat	55
Cardigan	47
Shirt	42
Gloves	38
Scarf	35
Hat	33
Socks	30
Furniture	
Bookcase	34
Sofa	31
Desk	28
Bed	24
Wardrobe	22
Table	20
Cupboard	18
Chair	15
Bedside table	11
Stool	9
Lamp	6
Picture	4

ing (1) adults' family resemblance and typicality ratings for the same items, (2) children's family resemblance ratings and both adults' family resemblance and typicality ratings on the same materials, and (3) children's typicality ratings generated with the use of Bjorklund et al.'s (1983) procedure and adults' family resemblance and typicality ratings for the same items.

The prediction made in Experiment 3 was that typicality ratings and family resemblance ratings produced

Table 5
Pearson's Product-Moment Correlations Between Typicality Ratings
Based on Bjorklund et al.'s (1983) Procedure, Typicality Ratings Based on
Degree of Liking, and Family Resemblance Ratings: Experiment 2

	Category							
	Animals		Fruit		Clothes		Furniture	
	Typicality	Liking	Typicality	Liking	Typicality	Liking	Typicality	Liking
Family resemblance	.49	.43	.33	.44	.44	.43	.33	.42

Note—$N = 96$. All correlations failed to reach significance ($p < 1$).

by adults for the same category items would be highly correlated. Previous work (see Rosch & Mervis, 1975) suggesting that family resemblance scores of category items are good predictors of typicality ratings for these category items provided support for this prediction. It was assumed that if we found (1) a high correspondence between both adults' family resemblance and typicality ratings for the same items and children's and adults' family resemblance ratings on the same materials and (2) a low correspondence in children's ratings generated by Bjorklund et al. (1983) to both family resemblance and typicality ratings generated by adults for the same items, we should be able to gain a strong support for the claim that the family resemblance method is a good way of obtaining typicality judgments from children and adults.

Method

Subjects. One hundred and twenty university students served as subjects. The mean age was 20 years, 1 month (19,0–21,25, $SD = 0.69$). The students were selected from the University of Crete and were predominantly from middle-class homes.

Materials and Design. The items were those used in Experiments 1 and 2. They are illustrated in the Appendix (see Experiment 1). Each item was given a featural definition by 10 subjects. Each subject provided featural definitions for four items, one from each category. Items were randomly selected within their categories so that each subject who received an item received it with different items from the other three categories and received the items representing each category in a different order. The subjects were also asked to rate the items of each category with the use of a rating task similar to that employed by previous investigators (see Rosch & Mervis, 1975). The rating task was presented a week after the family resemblance task had been completed.

Procedure. For the family resemblance ratings, the procedure was similar to the one used in Experiment 2 to derive family resemblance ratings from children, with the exception that the subjects were asked to write their answers on sheets of paper. The subjects were examined in groups of 10. Each of the 48 items shown in the Appendix was printed at the top of a page. Items were chosen randomly within a category so that each subject who received an item received it with different items from the other three categories and received the items representing each category in a different order. All subjects were told to pretend that some people from outer space who knew nothing about Earth had come to visit their town, and that they could help by telling them everything they knew about the category they were interested in learning about. The following instructions were given to the subjects:

Pretend that there are two people from outer space who know nothing about the planet Earth. They ask you to help them understand our world by giving some information about items belonging to certain categories. The names of these items are written on the pages that follow. Please write down as many attributes as you think are characteristic of each item. While doing so, remember that the people who are requiring

this information know nothing about these objects. Please try to give them as much information as possible. You have three minutes for each page. When I say turn to the next page, you do the same until you are told to turn the page again.

For the typicality ratings, the subjects were asked to rate, on a 7-point scale, the extent to which a category item represented their idea of the meaning of the category to which it belonged. Following previous investigators (Rosch & Mervis, 1975), each of the 12 items of each category was printed on a separate page. Items were chosen randomly within a category so that each subject received the items of each category in a different order. The order of presentation of the categories was counterbalanced across subjects. The instructions given were as follows

This is a simple experiment to find out how good or representative an example of a category is of the meaning of its category. There are four pages following this one. Twelve members of a particular category are written on each page. You are to rate the extent to which an item represents your idea or image of the meaning of the category name on a 7-point rating scale. A 1 means that you feel the item is an extremely good example of your idea or image of what the category is. A 7 means you feel the item fits very poorly with the idea or image of the category. A 4 means you feel the item fits moderately well. Use the other numbers of the 7-point scale to indicate the designated intermediate judgments. You will have 2 minutes for each page. When I say turn to the next page, you do the same until you are told to turn the page again.

Results and Discussion

The number of attributes applied by subjects to the items of each category were scored on the basis of the procedure adopted in Experiment 2. The measure of family resemblance for an item was, therefore, the sum of the weighted scores of each of the attributes that had been listed for that item. The measure of typicality was the mean rating on a 7-point scale of the extent to which items fit subjects' idea or image of the meaning of their category name. Items in each category were ranked 1–12 on the basis of the family resemblance scores. Similarly, they were ranked 1–12 on the basis of prototypicality. Tables 6 and 7 illustrate the items of each category listed, in their typicality order, on the basis of the rating scale and their family resemblance scores, respectively.

To test the hypothesis that there is correspondence between adults' typicality ratings of category items obtained with the family resemblance scores and typicality ratings of the same items generated in terms of a rating procedure, correlations between the ranks of items on family resemblance and their ranks on typicality were performed for each of the four categories. These correlations were highly significant ($r = .92$ for animals; $r = .94$ for fruit; $r = .98$ for clothes; $r = .95$ for furniture; $p < .0001$ for each category). These findings demonstrate that the relationship between the degree of family resemblance and

Table 6 Category Items Listed in Order of Their Typicality Scores: Experiment 3	
Category	Typicality
Animals	
Lion	6.8
Wolf	6.5
Dog	5.4
Cat	4.8
Goat	4.4
Monkey	4.0
Cow	3.8
Chicken	3.5
Turtle	3.0
Pigeon	2.7
Bee	2.2
Snake	1.8
Fruit	
Pear	6.9
Apple	6.0
Grapes	5.7
Cherries	5.2
Strawberries	4.9
Orange	4.6
Apricot	4.4
Pomegranate	4.0
Mandarin	3.5
Fig	3.1
Banana	2.8
Pineapple	2.5
Clothes	
Blouse	5.9
Trousers	5.5
Anorak	5.0
Skirt	4.8
Dress	4.6
Cardigan	4.4
Coat	4.0
Shirt	3.7
Gloves	3.3
Scarf	2.8
Socks	2.2
Hat	1.5
Furniture	
Bookcase	5.2
Bed	4.8
Sofa	4.5
Table	4.0
Desk	3.7
Chair	3.3
Bedside table	3.0
Wardrobe	2.6
Cupboard	2.2
Stool	1.9
Lamp	1.6
Picture	1.4

These correlations for the typicality ratings were as follows: animals, .96; fruit, .91; clothes, .95; furniture, .88. All were significant ($p < .0001$ and $p < .001$, respectively). These findings clearly show the close correspondence of children's family resemblance ratings to both adult ratings.

Correlations were also calculated between the ranks of items obtained from children with the Bjorklund et al. (1983) method in Experiment 1 and the ranks of the same items produced by adults in the present study on the basis of the family resemblance scores. These correlations failed

Table 7 Category Items Listed in Order of Their Family Resemblance Scores: Experiment 3	
Category	Family Resemblance
Animals	
Dog	108
Lion	102
Goat	98
Wolf	95
Monkey	91
Cat	88
Cow	84
Turtle	80
Chicken	75
Pigeon	68
Snake	60
Bee	52
Fruit	
Apple	86
Pear	81
Grapes	78
Cherries	72
Strawberries	68
Pomegranate	64
Orange	60
Mandarin	57
Fig	54
Apricot	51
Banana	48
Pineapple	46
Clothes	
Blouse	84
Anorak	80
Trousers	76
Dress	71
Skirt	67
Coat	63
Cardigan	60
Shirt	55
Gloves	52
Scarf	48
Socks	42
Hat	39
Furniture	
Bookcase	45
Bed	43
Sofa	41
Table	39
Chair	36
Wardrobe	31
Bedside table	28
Desk	22
Cupboard	16
Stool	14
Lamp	12
Picture	10

category typicality is robust. Such results confirm the hypothesis that family resemblance scores are good predictors of typicality ratings for adults.

Spearman rank-order correlations were also performed between the ranks of items on family resemblance generated by children in Experiment 2 and their ranks produced by adults in the present study in terms of family resemblance ratings and typicality ratings. These correlations, for the family resemblance ratings, were as follows: animals, .91; fruit, .97; furniture, .82; clothes, .96.

to reach significance ($r = .45$ for animals; $r = .48$ for fruit; $r = .43$ for clothes; $r = .45$ for furniture). Furthermore, the correlations performed between children's ratings derived according to Bjorklund et al.'s technique in Experiment 1 and adults' typicality ratings on the same materials were not significant ($r = .49$ for animals; $r = .50$ for fruit; $r = .48$ for clothes; $r = .50$ for furniture). Such results indicate that there is little correspondence between children's typicality ratings based on Bjorklund et al.'s method and adults' family resemblance ratings and typicality ratings for the same category items.

In summary, the preceding results revealed that adults' family resemblance and typicality ratings correlated highly, thus suggesting that the family resemblance method can be reliably used for derivation of typicality ratings from adults. The findings also demonstrated that both family resemblance and typicality ratings produced by adults were highly associated with children's family resemblance ratings and not significantly linked with children's ratings generated on the basis of Bjorklund et al.'s (1983) procedure from the same materials.

GENERAL DISCUSSION

The data presented here establish a close relationship between young children's typicality ratings derived on the basis of Bjorklund et al.'s (1983) technique and ratings produced in terms of an affective criterion. Five-year-old children failed to generate typicality judgments, despite the fact that the rating task was simplified and the instructions were given in a playful context. This pattern of results indicates that rating procedures cannot be reliably used to derive typicality judgments from children.

Three aspects of the findings, however, lend strong support to the claim that it is possible to establish typicality ratings from children with the use of the family resemblance method. First, adults' family resemblance ratings were found to correlate highly with their typicality ratings for the same materials. Such a result is consistent with previous findings (see Rosch & Mervis, 1975), suggesting that the family resemblance method is a reliable way of eliciting ratings from adults. Second, there was a robust relation between children's family resemblance ratings and adults' family resemblance and typicality ratings on the same materials and a low correlation between children's ratings generated with the use of Bjorklund et al.'s method (1983) and adults' family resemblance and typicality ratings for the same category items. Third, there was a low correlation between children's ratings obtained on the basis of Bjorklund et al.'s technique and their ratings derived in terms of the family resemblance method for the same materials.

Taken together, the data from this study suggest that not only are the family resemblance judgments reliable measures for deriving typicality ratings, but they are better measures of typicality than judgments based on a rating task proposed by Bjorklund et al. (1983) for young children. With respect to the available evidence, the conclusion may therefore be reached that the family resemblance method is a better way of obtaining typicality ratings from children than rating procedures are.

The reported data favor the view that the children could not respond to the linguistic and cognitive demands of the rating task. Thus, they were rating category exemplars as "good" and "poor" on the basis of their personal preferences rather than as a function of how representative they considered the items to be of their superordinate category. Further research, however, needs to be carried out to explore the influence of cognitive and linguistic factors on children's performance in rating procedures.

The family resemblance method of deriving ratings will be useful for researchers who wish to investigate children's performance in a variety of cognitive tasks, such as encoding of information into memory, category membership verification, processing of categorical information, and so forth (see Bjorklund, Smith, & Ornstein, 1982; Lin, Schwanenflugel, & Wisenbaker, 1991) with the use of child-generated typicality ratings.

REFERENCES

BJORKLUND, D. F., SMITH, S. C., & ORNSTEIN, P. A. (1982). Young children's release from proactive interference: The effects of category typicality. *Bulletin of the Psychonomic Society*, 20, 211-213.

BJORKLUND, D. F., THOMPSON, B. E., & ORNSTEIN, P. A. (1983). Developmental trends in children's typicality judgments. *Behavior Research Methods & Instrumentation*, 15, 350-356.

GLEITMAN, L. R., ARMSTRONG, L. S., & GLEITMAN, H. (1983). On doubting the concept "concept." In E. K. Scholnick (Ed.), *New trends in conceptual representations: Challenges to Piaget's theory?* (pp. 87-110). Hillsdale, NJ: Erlbaum.

HAMPTON, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, 18, 441-461.

LIN, P.-J., SCHWANENFLUGEL, P. J., & WISENBAKER, J. M. (1991). Category typicality, cultural familiarity, and the development of category knowledge. *Developmental Psychology*, 26, 805-813.

MARIDAKI-KASSOTAKI, K. (1993). *Conceptual instability in children and adults*. Unpublished doctoral dissertation, University of Lancaster.

ROSCH, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.

ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

APPENDIX
Category Items Listed in Order of Their
Production Frequency Scores: Experiment 1

Category	Production Frequency
Animals	
Cat	39
Dog	37
Cow	31
Chicken	27
Turtle	23
Pigeon	20
Goat	17
Monkey	15
Lion	12
Snake	8
Wolf	6
Bee	4
Fruit	
Apple	36
Orange	31
Pear	28
Banana	26
Strawberries	23
Cherries	21
Mandarin	18
Grapes	15
Apricot	13
Pomegranate	11
Fig	8
Pineapple	5
Clothes	
Cardigan	33
Trousers	28
Blouse	24
Shirt	21
Skirt	18
Anorak	16
Coat	14
Dress	11
Gloves	9
Scarf	7
Socks	5
Hat	3
Furniture	
Desk	36
Sofa	32
Bed	29
Bedside table	25
Stool	23
Table	21
Chair	18
Wardrobe	15
Bookcase	13
Picture	11
Cupboard	9
Lamp	4