The role of definitions in categorization and similarity judgments.

Serge Larochelle, Denis Cousineau and Annie Archambault

Université de Montréal

Abstract

Three experiments are reported that examined the role of definitions in categorization and similarity judgments. On each trial of the experiments, a verbal description was presented to the participants, along with a category name. Four different types of descriptions were used: Some asserted a set of properties deemed defining of the category named, while others asserted characteristic properties. Some negated a set of properties deemed necessary for category membership while others negated properties that were only characteristic of the category named. The results showed that necessary and characteristic properties are processed alike in both categorization and similarity judgments. Definitions and characteristic descriptions are also treated alike when making similarity judgments. However, definitions appear to be treated in a more unitary fashion when making categorization judgments. These results, which do not fit any of the main views of concepts (classical, probabilistic or dual), are explained in terms of the “theory” theory.

The research presented here bears on a venerable issue in cognitive science, namely: whether concepts have defining properties or not. One methodology that has been used to experimentally address this question consists in asking participants to determine whether some entity belongs or does not belong to a specified category on the basis of a description of the entity. In some cases, the description contains information that is deemed defining of the category, whereas in other cases, the descriptions lack such information. The present study is based on a similar approach. One major novelty is that the amount of information in the descriptions also varies. The combination of these two factors, i.e., description length and description type, provides a fairly strong rationale to determine whether concepts are well-defined or not. This rationale, like much research on categorization, is rooted in the classical view of concepts.

According to the classical view of concepts (Smith & Medin, 1981; Komatsu, 1992; Murphy, 2002), the mental representation of a category consists of properties that are common and unique to the objects in the category. As an example, the conceptual representation of the category labeled *square* would involve properties such as having four sides of equal length and four right angles. Since all squares are assumed to possess the above properties, a figure lacking any one of these properties cannot be a member of the category *square*. Each of the above properties is therefore said to be individually necessary for category membership. Since squares are the only figures to have all of the above properties, any figure with these properties must belong to the category *square*. The above properties are therefore jointly sufficient for category membership. Sets of individually necessary and jointly sufficient properties are considered defining because they allow to determine category membership in an all-or-none fashion. The rationale of the present research is based on the obvious, though unexploited implication that the number of properties in such defining sets should not affect categorization behavior. Whether a definition involves one, five or ten properties does not matter since objects need to have all the defining properties to belong to the designated category. Similarly, if an object lacks one, five or ten necessary properties does not matter either since the absence of a single necessary property is sufficient to reject category membership. Categorization behavior should also remain insensitive to the amount of characteristic information provided about an entity, so long as this information is insufficient to determine category membership.

The above predictions are graphically illustrated in Figure 1. The abscissa represents the number of properties making up various types of descriptions. The ordinate shows the level of confidence that the described entity belongs or does not belong to a specified category, depending on the nature and size of the description involved. The line labeled +Xd represents the performance that one would expect under the classical view when a set of X defining properties is asserted (+) about the entity to categorize. In this case, one should be absolutely confident that the entity described belongs to the specified category, irrespective of the number of properties in the definition. The line labeled –Xn represents the performance that one would expect when X necessary properties are negated (–) about the entity to categorize. Figure 1 shows no effect of description size since the lack of only one necessary property should be sufficient to reject category membership with absolute certainty. Finally, the lines labeled +Xc and –Xc represent descriptions that either assert or negate X characteristic properties. Since these properties are neither necessary nor sufficient for category membership, they do not allow any absolute decision about the nature of the entity described (The two lines in Figure 1 have been separated only for illustrative purpose).[1]

Insert Figure 1 about here

Very different predictions are obtained from a probabilistic view of conceptual representation. According to this view, the mental representation of categories involves properties that are common to some but not necessarily all members of a category. Such characteristic properties are also assumed to be involved in the categorization process. Although some probabilistic models have denied the existence of necessary and defining properties, such a strong stance is not essential (see Hampton, 1998). What is crucial is that all types of properties, whether necessary, defining or merely characteristic, be processed in the same way. While denying a special status for defining properties, the probabilistic view admits that some properties are more diagnostic than others, and are therefore weighed more heavily in categorical decisions. This is the case in models of the type described in Smith and Medin (1981) or proposed by Hampton (1979, 1998), for instance. Under a probabilistic view, categorization should therefore depend on the number and the diagnosticity of the properties asserted or negated about an object. Predictions derived from a probabilistic view are shown in Figure 2. As shown, confidence in categorical decisions is now graded. Confidence increases with the number of properties that are asserted about the entity to categorize and it decreases with the number of properties that are negated. The increase is more pronounced for +Xd than for +Xc descriptions because properties that are common and unique to members of a category are more diagnostic of category membership and presumably weigh more heavily in categorical decisions than properties that are merely characteristic. A similar reasoning applies to the performance expected with –Xn descriptions, compared to –Xc descriptions. Since the former, but not the latter, are made of properties that are thought to be common to all members of a category, they may weigh more heavily in categorical judgments, thereby causing a steeper change in confidence level with the number of properties negated.

Insert Figure 2 about here

Figure 3 shows the performance that one could expect under a dual-view of categorization. The predictions illustrated for +Xd and –Xn descriptions are identical to those in Figure 1, being based on the assumptions that defining descriptions are processed holistically and that the negation of a single necessary property is sufficient to reject category membership. The predictions illustrated for +Xc and –Xc descriptions are identical to those in Figure 2, being based on the assumption that characteristic properties individually contribute a certain weight to categorical decisions. Dual views of concept representation come under different guises. According to one proposal, characteristic properties are used mostly for the quick identification of objects whereas true categorization must rest on a conceptual core of defining properties (Osherson & Smith, 1981). To use a well-known example, the sight of a cookie-baking, gray-haired woman may lead to the hypothesis that the person is a grand-mother. However, for true categorization, one must test whether the person is the mother of a parent. The instructions given to the participants in the experiments reported here emphasized true categorization.

Insert Figure 3 about here

A critical aspect of dual views is that defining properties be distinct from non-defining properties, so as to be processed differently. This was the case in Smith, Shoben and Rips's (1974) early model of category verification times. Defining and non-defining properties are also processed differently in McNamara and Sternberg's (1983) mixed model. In the model, one mechanism checks for the presence of necessary or defining properties and a second one accumulates the evidence provided by characteristic properties. The latter type of property is assumed to determine categorization decisions when necessary or defining properties are absent, in which case the model's decision sums the weight of the evidence provided by the characteristic properties. Therefore McNamara & Sternberg's mixed model would make predictions very similar to those illustrated in Figure 3. Descriptions of the four types discussed so far (i.e., +Xd, +Xc, -Xn, and –Xc) are involved in Experiment 1.

Testing dual views is complicated by the fact that some theories do not require all concepts to have both defining and characteristic properties. This is the case of psychological essentialism (Medin & Ortony,1989), for instance. According to the theory, there is a deeply ingrained belief that objects have permanent, essential properties, which make them what they are, alongside more variable and accidental properties. Such a belief in essences is assumed to apply to all natural kinds. For some categories, the belief in essence may be accompanied by knowledge of properties that are essential, or so intrinsically tied to the essential properties as to form a definition in the classical sense while, for other categories, the same person may hold more or less "inchoate" ideas (Medin & Ortony, 1989, p. 184), or even rely on scientists to specify essential properties now or in the future (Putnam, 1973, 1975). In the context of dual-view theories, an attempt to distinguish defining from non-defining properties must focus on categories for which subjects can provide defining properties, the question being whether they treat such properties differently from non-defining ones.

Another complication is that there are many different ways to integrate the information provided by descriptions. The predictions illustrated in Figures 1 to 3 are based on the assumptions that the characteristic properties have an additive effect on confidence level and that the expected change in confidence with every additional property is fairly constant. An alternate way to integrate the information would be to average the weight of the properties in the descriptions. Averaging rules are prevalent in information integration theory (Anderson, 1981). In the field of categorization, such a rule has been proposed by Hampton (1988) to account for performance in experiments involving conjunctive concepts. In the present paradigm, performance would no longer depend on the number of properties in the description if the property weights were averaged instead of being summed. However, performance would still depend on the type of description. In short, the performance expected of probabilistic or mixed models based on an averaging rule could resemble that illustrated in Figure 1, thereby making the different views quite difficult to distinguish.

Such a distinction can be achieved, however, by incorporating conflicting information in the descriptions. Imagine a description asserting that an entity has all the defining properties of a specified category, but that it lacks a property that is characteristic of some members of the category. Such descriptions can be labeled +Xd-1c. According to the classical view, performance should not be affected by the absence of a characteristic property. The same would also be true of dual models based on the assumption that characteristic information is ignored when a definition is present. However, even if one were to allow conflicting characteristic information to be taken into account, the effect on performance should still be independent of the number of properties in the definition, since definitions are assumed to have a holistic effect in classical and dual views. So, although the confidence level obtained with +Xd-1c descriptions might be smaller than that obtained with +Xd descriptions, it should not vary with definition size. By contrast, in a probabilistic model based on an averaging rule, the negation of a characteristic property should have differential effects on performance depending on the number of properties in the definition. The effect of the negated property on the mean weight should decrease as the number of asserted properties increases and the denominator gets larger, thereby allowing confidence to increase with the number of defining properties in +Xd-1c descriptions. The same prediction applies to other types of descriptions containing conflicting evidence, such as +Xc-1c, -Xn+1c and –Xc+1c. Not only does a probabilistic model based on an averaging rule still predict a change in confidence with description size, but the change in confidence level should be more pronounced for +Xd-1c than for +Xc-1c, and for –Xn+1c than for –Xc+1c, given the larger weight presumably associated with necessary and defining properties. In short, with descriptions containing conflicting information, models based on a sum vs. an average of weights make a similar pattern of predictions, the main difference between models residing in the linear vs. non-linear shape of the description size functions. Descriptions involving conflicting information were used in Experiments 2 and 3, reported here.

Importance rating and property selection

One precondition for testing the hypotheses outlined in the Introduction is that participants be able to identify properties deemed necessary and defining of various categories. Another is that such properties be considered to have more weight than merely characteristic properties. These two prerequisites were tested in the first two sessions of each experiment. The aim of the first session was to obtain the participants' judgments concerning the importance of every property with respect to the associated category name. The goal of the second session, which had two parts, was to identify the properties that participants considered necessary (part one), and necessary and sufficient (part two). We will describe these two sessions, which were identical over all experiments, before turning to the results of the third session, which differed across experiments.

Method

Participants

The participants in this study were undergraduate students at the Université de Montréal, enrolled in Departments other than Linguistics, Philosophy or Psychology. Fourteen women and 8 men, aged between 19 and 31, participated in Experiment 1. Thirteen women and 3 men, aged between 19 and 23, participated in Experiment 2. Sixteen other students (11 women and 5 men) participated in Experiment 3. Their age ranged from 19 to 22. All were native speakers of French.

Stimuli

All the stimulus material was in French. The category names, which are listed in Table 1, were taken from an earlier study by Saumier (1993). Of the 24 category names listed, eight referred to artifacts, eight to biological kinds and eight were assigned to a group called “well-defined”.[2] A list of properties was associated to each category name. These property lists were also taken from Saumier (1993). They had been obtained by asking a group of undergraduate students to write down properties that describe the members of the specified category. No constraint had been placed on the necessity, sufficiency or even the importance of the properties that the participants could give. The properties gathered by Saumier were generally used verbatim. The properties that were modified involved an indefinite quantifier over the members of a category. Since it would be the participants' task to determine whether a property was true of all members of a category or not, we deleted all indefinite quantifiers from these properties, whether they were stated explicitly (e.g., "some are found in the countryside" for *sparrow*) or implicitly through modals (e.g., "may have armrests" for *chair*).

Insert Table 1 about here

Procedure

Participants did the first two sessions on separate days at least one week apart. Each session lasted between one and two hours. These sessions were run using a spreadsheet software with a separate file for each category. The category name appeared on top of the computer screen with each of its properties listed below on separate lines. The properties were listed either in an alphabetical order or in a reversed alphabetical order, depending on the participants and categories. Both orders were represented equally often over categories and participants.

Participants were told to interpret the category names literally rather than metaphorically, and to choose and stick to one meaning when they considered the category name to have more than one. For the first session, participants were instructed to rate the importance of each property with respect to the category named, using a scale going from 1 (extremely small importance) to 9 (extremely large importance). Participants typed their ratings next to the corresponding properties. Order of presentation of the various category names and associated property lists was mixed over category types and differed across participants.

Subjects saw the same category names and associated property lists in the same order in session 2 as they had in session 1. However, the importance ratings gathered in session 1 were hidden from the participants. The first part of session 2 was devoted to the identification of properties deemed necessary for category membership. Participants were asked to distinguish properties considered to be “common to all category members” (by giving a rating of 2 to such properties), from those thought to be "present in some but not all members" (which were given a rating of 1). Finally, a rating of 0 was to be given to properties considered to be "present in none of the category members." [3] Emphasis was placed on the fact that the properties listed could belong to members of other categories, but that this was irrelevant since the judgments required concerned only the category named.

The category names were once again presented in the same order in the second part of session 2 but, this time, the associated property lists were reduced to only those properties having previously been identified as necessary by the participant. All prior ratings were also hidden from sight. Participants were asked to find the smallest set of properties that was also deemed “unique” to members of the named category. Emphasis was placed on the fact that each property in this set could be shared by members of other categories (provided that the set contained more than one property), but that the set as a whole had to be "true only" of the category named. Participants were asked to attribute a rating of 1 to all properties in the defining set selected and a 0 to the remaining properties, attributing a 0 to all the properties listed if they did not find any defining set.

Results

As shown in Table 2, almost all participants identified at least one necessary property for almost all items. Ninety-five percent of the participants also succeeded in specifying a definition for the well-defined terms. The corresponding percentages were 86% for biological kinds and 92% for artifacts. Table 2 also gives the mean number of properties identified as necessary and defining for each type of category name involved in the Experiments.[4] As shown, fewer properties were selected as necessary for well-defined terms than for artifact and biological kind terms. Table 2 shows that the definitions produced by the participants for the well-defined terms also contained a smaller number of properties than those concerning the other category types. [5]

Insert Table 2 about here

These results show that participants are quite confident not only in the existence of necessary and defining properties for various category names, but also in their knowledge of such properties, which was a prerequisite for testing the main hypotheses of the present study. Similar, though not identical results, had previously been obtained by McNamara and Sternberg (1983). Their subjects, like ours, identified necessary properties for almost all the test items in their study. However, their participants identified defining properties for only about 58% of the natural kind terms and 45% of the artifact terms, compared to close to 90% for both category types in our Experiments. This difference is probably attributable to differences in procedure across the two studies. Participants in McNamara and Sternberg's (1983) study did not have to select defining properties. Instead, they first had to select necessary properties among those present on a master list, also containing characteristic properties. Then, they had to select a set of sufficient properties from the same master list. The properties deemed defining for a given item resulted from the overlap between these two successive choices. By contrast, our participants were explicitly asked to identify a set of defining properties by selecting a set of sufficient properties among those they had previously chosen as necessary. It may therefore have been easier for them to do so. Alternatively, one could argue that our procedure induced participants to identify sets of properties that were not always really thought to be sufficient. If this were the case, then categorization performance would be biased towards probabilistic predictions, which is contrary to the findings described shortly.

A second condition for testing the hypotheses stated in the Introduction is that necessary and defining properties be given more weight than characteristic ones. Although not presented in Table 2, the importance ratings of the characteristic properties were smaller than those of the necessary properties for all category names in all Experiments. On average over all three experiments, characteristic properties received an importance rating of 5.43 for well-defined terms, of 5.79 for biological kind terms and of 5.91 for artifact terms. In comparison, the average importance ratings for necessary properties were 7.33 for well-defined terms, 7.15 for biological kind terms and 7.24 for artifacts. As shown in Table 2, ratings were even higher for the necessary properties that were selected to form a defining set. In short, the second condition for testing the effects of description type on categorization behavior, namely that necessary and defining properties be weighed more heavily than characteristic properties, appears to be met.

Categorization judgments

Method

The third and critical session of Experiments 1 and 2 was devoted to a categorization task. Participants were told that they would have to identify entities on the sole basis of the descriptions provided. A rating of 9 was used to indicate "absolute certainty" that the entity described was a member of the category named. At the opposite, a rating of 1 would indicate "absolute certainty" that the described entity was not a member of the category named. Subjects were told that they did not have to identify the described entity at all costs. Emphasis was placed instead on the importance of giving the answer most logical or most likely to be true. A rating of 5 was to be given if the description did not allow to decide whether the entity could or could not bear the name presented. Ratings between 5 and 9, and between 5 and 1, indicated an increasing certainty that the entity described belonged or did not belong to the category named.

Stimuli

The descriptions used in Experiments 1 and 2 were each individually tailored to each participant and category name. Experiment 1 involved four different types of descriptions. +Xd descriptions were composed of the properties having previously been identified by the participant (part 2 of session 2) as forming a classical definition for the corresponding category name. –Xn descriptions were composed of the negative version of properties having previously been judged by the participant (part 1 of session 2) as being true of all members of the category.[6] The necessary properties in –Xn descriptions were chosen randomly, the number of properties in such descriptions being the same as in the corresponding +Xd descriptions. In +Xc descriptions, a set of characteristic properties was asserted. These properties were randomly selected among those thought by the participant to be true of only some members of the category. Finally, –Xc descriptions were formed of the negative version of randomly selected characteristic properties. The number of properties in +Xc and –Xc descriptions was the same as in the corresponding +Xd and -Xn descriptions for a given participant and category name.

As mentioned in the Introduction, distinguishing among views of categorization can be facilitated by introducing conflicting information in the descriptions. This was done in Experiment 2, which involved eight different types of descriptions. Two types of descriptions contained the properties considered to be defining of a given category name by the participant, along with one or four properties thought to be only characteristic of the category. The defining properties were asserted and the characteristic properties were negated, thus the labels +Xd-1c and +Xd-4c. In two other types of descriptions, labeled

+Xc-1c and +Xc-4c, different sets of characteristic properties were asserted and negated. The number of properties asserted was smaller or equal to the number of properties deemed defining of the same category by the same participant, and one or four other characteristic properties were negated. There were two description types, labeled -Xn+1c and -Xn+4c, in which a number of necessary properties, equal to the number of properties in the corresponding definition, were negated and one or four characteristic properties were asserted. Finally, in the last two descriptions, labeled -Xc+1c and -Xc+4c, a number of characteristic properties, smaller or equal to the number of properties in the definition, were negated and one or four other characteristic properties were asserted.

Apart from the theoretical motivation sketched in the Introduction, the use of descriptions containing conflicting information in Experiment 2 served a useful methodological purpose: They reduce the likelihood that flat description size functions in Experiment 1 result from ceiling (in the case of +Xd descriptions) or floor effects (in the case of –Xn descriptions). By moving the ratings away from the boundaries of the rating scale, the presence of conflicting characteristic information could make room for the ratings to increase with the number of defining properties and decrease with the number of necessary properties. Failure to find such changes in ratings with description size in Experiment 2 would constitute further evidence for classical or dual views, and so would the finding that confidence levels are not affected by the number of conflicting characteristic properties.

In both experiments, descriptions were made only for category names having been assigned a definition by the participant. Even for such items, it was not always possible to construct all the required types of descriptions. In some cases, there were too few necessary properties left to negate. In others, there were too few or no characteristic properties to assert or to negate. Participants were tested with items for which all the matched description types could be made. In Experiment 1, each participant was also tested with the descriptions constructed for another, randomly selected subject. These other descriptions were intended as fillers. Participants in Experiment 2 saw only descriptions derived from their own property selection protocol.

Procedure

The third session was done a week to a month after session two and lasted between 90 and 120 mn. Testing order of the various category names and description types was randomized in both experiments. On each trial of Experiment 1, a description first appeared on top of the screen, one property per line in random order. Participants were instructed to read the description and then press the space bar. This caused a category name to appear below the description, along with the rating scale. Participants were warned that their task was not to guess the category name that would appear. Rather, the task was to give the most logical response considering the description and category name presented. Participants indicated the chosen rating by typing the corresponding digit on the keyboard. In Experiment 2, the descriptions and category names were displayed simultaneously on the screen, along with the rating scale. Asserted and negated properties were randomly mixed on the screen.

Results

Experiment 1

A total of 1768 descriptions were made for the 22 participants in Experiment 1. Since construction of these descriptions was not fully automated, errors were unavoidable. All descriptions found to contain an error were excluded from the analyses. Also excluded were the descriptions that matched a description containing an error. For instance, if an error was found in the +Xd description of a given participant for a given category name, the matching +Xc description was also excluded from the data, and vice versa. The same was done for –Xn and –Xc descriptions. This caused the loss of about 2.0% of the data, a total of 1736 descriptions remaining over all participants, category names and description types. In short, the data used to contrast performance obtained with descriptions containing only characteristic properties versus those containing defining or necessary properties came from the same participant-category pairs.

Since the size of +Xd descriptions (i.e., definition size) had been determined by the participants in part 2 of session 2 and, since it determined the size of the matching +Xc, -Xn and –Xc descriptions, there was little chance that all conditions of the Experiment be equally represented under all description sizes. Over all participants and category names, description size varied from 1 to over 10 propositions. [7] Not all category types were represented among descriptions of size 8 or more. Even for description sizes smaller than 8, there was no guarantee that all category names would contribute observations, much less an even number of observations. This uneven distribution of observations over description size led us to compute a separate linear regression for every category name and description type, using the number of propositions in the descriptions as predictor variable and the categorization ratings provided by different participants as predicted variable. The slopes of these regressions, as well as other measures, were then entered into ANOVAs, using category name as the random factor. Category type was considered a “between” factor, whereas description type was considered a “within” factor. Separate ANOVAS were computed on the results obtained with positive descriptions (+Xd vs. +Xc) and with negative descriptions (-Xn and –Xc). Note that our use of regression coefficients as a dependent variable in the ANOVAs does not imply that the description size effect is strictly linear, or even that it has a significant linear component. Slopes are simply used as a measure of the description size effects, one that takes into account all the available data without requiring that every description size be equally represented within each category name.

Insert Figure 4 about here

Figure 4 summarizes the results obtained in Experiment 1. A separate panel is devoted to each category type. The height of the four points in each panel corresponds to the mean rating, computed over category names, for each description type. The location of the points along the abscissa corresponds to the average number of propositions. The four lines in each panel were drawn using the average intercepts and slopes computed over category names for each description type. The main result obtained with positive descriptions is that slopes were generally smaller for defining descriptions (mean = 0.08) than for descriptions made of characteristic properties (mean = 0.21). This convergence of the regression lines for +Xd and +Xc descriptions did not differ significantly across category types. Turning to negative descriptions, Figure 4 shows some convergence of the regression lines for -Xn and -Xc descriptions. However, the statistical analyses did not reveal the slopes obtained with –Xn (mean = - 0.12) and –Xc (mean = - 0.19) to differ from each other for any category type.

The mean rating obtained with +Xd descriptions (8.29) was larger than that obtained with +Xc descriptions (6.19). The difference in mean rating between these description types did not vary significantly across category types. The same was true of the difference in mean rating between –Xn and –Xc descriptions. However, in the case of negative descriptions, the mean rating obtained with descriptions containing necessary properties (1.94) was smaller than that obtained with descriptions made of characteristic properties (2.79). An interesting finding is that the mean ratings obtained with –Xc descriptions were about twice as far from the center of the scale (5) as those obtained with +Xc descriptions. This result suggests that negating characteristic properties about an entity has more influence on categorization than asserting similar properties or, perhaps more generally, that positive and negative evidence are weighed differently in categorical decision.

Although the pattern of results obtained in Experiment 1 and shown in Figure 4 appears quite similar to that predicted by some dual models, especially in the case of well-defined terms, the conclusions that are statistically supported are the following: First, the description size effects obtained with +Xc and –Xc descriptions show that characteristic information is taken into account when no other information is available for categorization (as just about everyone would have expected). Second, the lack of difference in slopes between –Xn and –Xc descriptions shows that necessary properties may not be processed differently from characteristic properties, although the difference in mean ratings shows that necessary properties weigh more heavily in categorical decisions. The third finding, which is somewhat surprising given the previous one, is that defining properties not only appear to weigh more than characteristic properties, as evidenced by the larger average ratings obtained with +Xd descriptions, but they also appear to be processed in a more unitary fashion, as indicated by the shallower slopes obtained with +Xd descriptions. There was no reliable evidence that participants behaved differently depending on the type of category involved.

As mentioned earlier, the results of Experiment 1 could have been contaminated by ceiling and floor effects of a methodological nature. This would have been especially likely to occur for +Xd and –Xn descriptions since the mean ratings obtained with such descriptions were quite close to the extremities of the rating scale. Experiment 2 was less sensitive to this possible bias.

Experiment 2

As in Experiment 1, the data used to contrast performance obtained with descriptions containing only characteristic properties versus those also containing defining or necessary properties came from the same participant-category pairs. This was achieved by eliminating every description containing an error along with the matched description (e.g., +Xd-1c and +Xc-1c). Such error elimination left a total of 2482 trials, for an average of over 150 per participant. To ease comparison with the results of Experiment 1, description size was measured by counting the propositions forming the variable part of the descriptions (i.e., the +Xd, -Xn and +/-Xc parts) and ignoring the propositions forming the constant parts (i.e., the +/-1c and +/-4c parts). We will present the results obtained with one and four conflicting characteristic properties together, noting differences when they occur. We will call “positive” the descriptions that asserted a variable number of defining (+Xd) or characteristic properties (+Xc) and drop the –1c and -4c from the label of the conditions, except when needed. Similarly, the descriptions that negated a variable number of necessary or characteristic properties will be called “negative” and labeled –Xn and –Xc, despite the assertion of one or four characteristic properties.

Insert Figure 5 about here

Figure 5 summarizes the results in Experiment 2, averaged over both conflict conditions. The format of the Figure is the same as that of Figure 4. The most obvious result is that performance was affected by the presence of conflicting characteristic information. The mean rating obtained with +Xd-4c descriptions (5.11) was smaller than that obtained with +Xd-1c descriptions (7.20), which was in turn smaller than that obtained with +Xd descriptions in Experiment 1 (8.29). This result shows that characteristic properties are taken into account, even in the presence of defining properties. Descriptions made of characteristic properties were also affected by conflicting information. The mean rating obtained with +Xc-4c descriptions (3.21) was smaller than that obtained with +Xc-1c descriptions (5.01), which was in turn smaller than that obtained with +Xc descriptions in Experiment 1 (6.19). The overall difference in mean rating between +Xd descriptions (6.16) and +Xc descriptions (4.11) was significant in Experiment 2, as it was in Experiment 1. Category type did not have any effect on mean ratings, nor did it interact with any other factor in the analysis.

Given that the mean ratings obtained with positive descriptions in Experiment 2 were smaller than those obtained in Experiment 1, the slopes should be less affected by ceiling effects. The average slope obtained with +Xd descriptions in Experiment 2 (0.22) was indeed larger than that obtained in Experiment 1 (0.08). Nonetheless, the slopes for +Xd descriptions in Experiment 2 remained smaller, on average, than those obtained with +Xc descriptions (0.45). Description type did not interact with the number of conflicting properties, the mean slopes obtained with one or four conflicting properties being very similar for +Xd descriptions (0.23 and 0.22, respectively) and for +Xc descriptions (0.48 and 0.42, respectively). Figure 5 shows the convergence of the regression lines for +Xd and +Xc descriptions to be less pronounced for artifacts than for the other category types, which resulted in a marginally significant Description X Category type interaction.

Turning to negative descriptions, those containing necessary properties yielded a smaller average rating (2.79) than descriptions composed only of characteristic properties (4.00). Both types of descriptions were affected by the amount of conflicting information. The mean rating obtained with –Xc+4c descriptions (5.15) was larger than that obtained with –Xc+1c descriptions (2.85), the latter being similar to the rating obtained with -Xc descriptions in Experiment 1 (2.79). Similarly, the mean rating obtained with –Xn+4c (3.64) descriptions was larger than that obtained with –Xn+1c descriptions (1.94), but the latter was equal to the rating obtained with -Xn descriptions in Experiment 1. Given the small ratings obtained, floor effects were still possible with –Xn+1c descriptions in Experiment 2. Category type had a close to significant effect on mean ratings, due to the somewhat larger ratings obtained with well-defined terms. However, category type failed to interact with any other factor in the analysis.

As in Experiment 1, the mean slope obtained with –Xn descriptions (-0.13) did not differ significantly from that obtained with –Xc descriptions (-0.17). However, the slopes for –Xn descriptions were affected by the number of conflicting properties, being smaller with one conflicting property (-0.05) than with four (-0.22). As mentioned, the small slopes obtained with –Xn+1c descriptions could still be due to a floor effect. This possibility is supported by the fact that the slopes for –Xc+1c (-0.16) and –Xc+4c (-0.17) were almost identical and comparable to that obtained with –Xn+4c descriptions. In short, apart from some possible floor effects, there was no evidence of any convergence in the regression lines obtained with –Xn and –Xc descriptions for any of the category types.

Finally, it may be worth noting that negative descriptions generally produced more extreme ratings than positive descriptions, which suggests once again that positive and negative evidence are not weighed equally in categorical decisions.

Discussion

The results of Experiment 2 are generally in line with those of Experiment 1. However, Experiment 2 extends the findings of Experiment 1 by showing that characteristic information is taken into account in categorization judgments, even when necessary or defining information is present. Such a result is clearly at odds with a strict classical view. The results of both experiments also conflict with a purely probabilistic view. Probabilistic models incorporating a summing rule predict more pronounced description size effects for +Xd and –Xn descriptions than for +Xc and –Xc descriptions, respectively.[8] Probabilistic models based on an averaging rule make similar predictions for description incorporating conflicting information, although the differences in slopes would not be pronounced as in a summing model. In fact, none of the category types produced larger description size effects for descriptions containing necessary or defining properties than for descriptions made only of characteristic properties.

The results fit best, though not perfectly, a dual view of concept representation. Although the slopes obtained with +Xd descriptions in Experiment 2 were larger than zero, they were still much smaller than those obtained with +Xc descriptions, which suggests that definitions were processed in a more holistic, though not perfectly unitary fashion. Since the analyses reported here concerned only items for which the subjects could provide a definition, the differential size effects obtained with +Xd and +Xc properties suggest that participants did distinguish between defining and non-defining properties in the representation of the same items, which is the essence of a dual view. A dual view can also accommodate the effects obtained with characteristic properties inasmuch as it is assumed that such information is taken into account, in the presence as well as in the absence of necessary or defining information. The main difficulty for a dual view is that -Xn and -Xc properties did not seem to be processed differently, both conditions exhibiting similar description size effects. Smith and Sloman (1994) suggested that rule-based categorization might be restricted to situations where no characteristic information is available. Since such information was present in the descriptions of Experiment 2, their proposal is consistent with the lack of convergence obtained with –Xn and –Xc descriptions but, by the same token, it fails to account for the convergence obtained with +Xd and +Xc descriptions. The results obtained with –Xn description in Experiments 1 and 2 are also inconsistent with McNamara and Sternberg's (1983) mixed model.

Murphy and Medin's (1985) "theory" theory could possibly account for the pattern of results obtained. Their approach rejects the notion that concepts are represented by lists of isolated features or properties. Instead, the properties of a concept are thought to be interconnected by a complex web of relations, foremost among which are causal relations, that provide an explanatory theory of what it takes to be a member of a category. As Pothos and Hahn (2000) have suggested, such explanatory theory can serve as the basis for selecting properties that are considered necessary and/or sufficient for category membership. Suppose that this was the case in our Experiments. By virtue of being based on an underlying theory, the properties selected as defining would have been more critical for category membership than properties considered to be only characteristic. So, defining properties would have had more weight than characteristic properties. Moreover, since the entire set of properties considered defining by the participants were included in +Xd descriptions, these descriptions would have preserved whatever relationships the participants found to exist among these properties. In other words, the descriptions made of defining properties would have maintained the coherence of the underlying theory. By contrast, the descriptions made of characteristic properties would have had little conceptual cohesion since the constituting properties were selected randomly. This difference in cohesiveness would explain why descriptions made of defining properties were treated in a more holistic fashion than descriptions made of characteristic properties. With +Xc descriptions, lack of conceptual cohesion may have left the participants little choice but to process each property separately. The properties selected as necessary by the participants were also presumably more critical for category membership than those selected as characteristic. However, the necessary properties entering in -Xn descriptions were selected randomly. More importantly, they were negated. As a result, -Xn descriptions could only violate the theories of what it takes to be a member of a category, which would explain why such descriptions were processed in the same piecemeal fashion as –Xc descriptions.

The main evidence quoted in favor of the “theory” theory comes from dissociations between categorization and similarity judgments (Rips, 1989; Rips & Collins, 1993), the idea being that the former, but not the latter, rely on a theory of what it takes to be a member of a category. A typical example, mentioned in the Introduction, is that of a gray-haired woman, who may look very much like a grand-mother, but who cannot be one if she is not the mother of a parent, because this is essential to the theory of grand-motherhood. If similarity judgments need not rely on a theory, then the descriptions containing defining properties should be processed in the same way as descriptions made of characteristic properties. In other words, the description size effects obtained with +Xd descriptions should be similar to those obtained with +Xc descriptions. Testing this hypothesis was the goal of Experiment 3.

Similarity judgments

Method

In Experiment 3, subjects had to judge how similar the described entity was to members of the category named. The ratings were done on a nine point scale, from “extremely dissimilar” (rating of 1) to “extremely similar” (rating of 9). Using each individual participant’s property selection protocol, we constructed the same eight different types of descriptions as in Experiment 2. Except for the nature of the judgment required, the procedure was identical to that of Experiment 2. Error elimination left a total of 2532 trials.

Results

Figure 6 shows the mean ratings, averaged over both conflict conditions, along with the best fitting average regression lines. The main new finding of Experiment 3 is that description type failed to have a significant effect on the slopes obtained with positive descriptions, the mean slopes being 0.31 and 0.37 for +Xd and +Xc descriptions, respectively. The interaction between description and category type failed to reach significance, and so did the main category type effect. In short, similarity judgments seem to have been affected by the size of definitions about as much as they were by the size of descriptions containing only characteristic descriptions. The number of conflicting properties had no effect on the slopes, nor did it interact with category type or description type. To test the reliability of differences between categorization and similarity judgments, we compared the slopes obtained in Experiments 2 and 3. There was no overall difference across experiments in the slopes obtained with positive descriptions but judgment type did interact with description type, this interaction being the sole significant one in the ANOVA (all other ps > .10).

Insert Figure 6 about here

The mean ratings obtained in Experiment 3 were affected by the amount of conflicting information in the descriptions. The mean similarity rating obtained with +Xd-1c descriptions (7.15) was larger than that obtained with +Xd-4c descriptions (4.88), and so was the mean rating obtained with +Xc-1c descriptions (5.38), compared to that obtained with +Xc-4c descriptions (3.31). The overall difference between +Xd descriptions (6.01) and +Xc descriptions (4.34) was significant. Compared to the categorization ratings obtained in Experiment 2, the similarity ratings obtained with +Xd descriptions were slightly smaller on average (6.02 vs. 6.16), while those obtained with +Xc descriptions were slightly larger (4.34 vs. 4.11), which resulted in a significant interaction between judgment and description type. There was no other interaction involving judgment type in the analysis comparing the ratings obtained with positive descriptions in Experiments 2 and 3.

The mean ratings obtained with negative descriptions were more extreme than those obtained with positive descriptions, as usual. However, they were also affected by the number of conflicting properties. The mean similarity rating obtained with -Xn+1c descriptions was 2.17, compared to 4.44 for –Xn+4c. The mean similarity ratings for –Xc+1c and –Xc+4c were 3.08 and 5.52, respectively. The presence of a 3-way interaction involving category type indicates that there were variations across categories in the relative resistance of –Xn and –Xc descriptions to the amount of conflicting information. Despite this interaction, -Xn descriptions always produced smaller ratings (mean = 3.31) than –Xc descriptions (mean = 4.30). Comparing the similarity ratings of Experiment 3 with the categorization ratings of Experiment 2 revealed that the former were less extreme. The differences in mean ratings across experiments were more pronounced with descriptions containing four conflicting property, than with descriptions containing only one such property. This was especially true of –Xn descriptions, so that performance obtained with –Xn+1c could still be affected by a floor effect.

As shown in Figure 6, the slopes obtained with –Xn descriptions did not differ, on average, from those obtained with –Xc descriptions. However, these mean slopes hide an interaction with number of conflicting properties, similar to that found in Experiment 2. The slopes obtained with –Xn descriptions were affected by the number of conflicting properties, the mean slope for –Xn+1c descriptions being -0.12 versus -0.42 for –Xn+4c descriptions. There was little variation in this pattern of effect over categories. The small slopes obtained with –Xn+1c descriptions could still reflect some floor effect, given that the mean rating was still very low. By contrast, the slopes obtained with –Xc+1c (mean = -0.26) and –Xc+4c (mean = -0.24) descriptions were almost identical, but there were some variations in the size of the difference across category types. These variations caused a significant 3-way interaction involving category and description type along with number of conflicting properties. Since there was no such interaction in the slopes obtained with negative descriptions in Experiment 2, the ANOVA comparing the results of both experiments showed a significant 4-way interaction.

General Discussion

Let us first briefly review the main findings of this study, in deceasing order of certainty. The first finding is that the weight of information on categorization and similarity judgments, depends on whether the properties are asserted or negated about the entity concerned. Negated properties, whether necessary or merely characteristic, had a larger influence on the average categorization and similarity ratings, than properties that were asserted, be they defining of characteristic. The second finding is that, although some properties are identified as common to all category members, they are not treated as necessary for category membership. Hampton (1995) made valiant efforts to construct descriptions comprising properties that would be considered necessary for category membership by the participants in the study. His repeated failures, along with the present findings, suggest that no property is considered as such. Rips' (1989) study is often quoted as evidence against this conclusion, but Rips' results were obtained in very limited conditions and, even in such conditions, it has been difficult to replicate them (see Smith & Sloman, 1994; Pothos & Hahn, 2000).

The third finding is that definitions seem to have a special status in categorization. This finding is perplexing from a classical perspective: How can definitions receive special treatment when the necessary properties that compose them do not? One possible explanation of this result is that people rely more on sufficiency than on necessity in categorization. The problem with such an interpretation is it requires two different notions of sufficiency: Why would the assertion of a set of unique properties be “sufficient” to accept category membership while the negation of common properties would not be “sufficient” to reject category membership? This explanation is also problematic from a computational point of view, since testing necessity seems less costly than testing sufficiency. In terms of processing, it can certainly be arduous to determine whether a given property applies to all members of a category when the category is large but, at least, it is a finite task. By contrast, determining whether a set of properties is unique to the members of a category requires to consider an indefinite number of alternative categories, which can be an endless task unless the decision can be reached deductively instead of inductively. For this reason, we prefer to attribute the results obtained with definitions to their conceptual coherence. If this interpretation is correct, then explanatory coherence does not seem to play the same role in similarity judgment. The fact that the similarity ratings were not as affected by the nature of the properties in the descriptions is a fourth important finding of the present study.

We have argued that descriptions containing necessary properties lacked conceptual coherence since these properties were negated about the entities to categorize. Alternately, one could argue that these descriptions lacked cohesion because the necessary properties, like the characteristic properties, were selected randomly. An easy way to discriminate these possibilities would be to negate all the necessary properties in the definitions. Such descriptions would have as much surface cohesion as the corresponding definitions, but they would lack the underlying conceptual coherence. Note that the outcome of such a test would not invalidate our previous conclusion about the lack of necessary properties. A set of truly necessary properties should still be treated as such, even when they are selected randomly.

The results reported so far concern the effects of description type and size on categorization and similarity judgments. They are based on a rather indirect measure, namely regression coefficients. Our interest in description size stemmed from the fact that different theories and models of categorization make different predictions about its effect on performance. However, the critical aspect of these theories and models does not lie in description size per se, but in the nature of the rules used to integrate the available information. Regression coefficients allowed to test some of the models' predictions, but they prohibited addressing others, namely those having to do with the rule(s) used to combine the information provided by the descriptions. Since many models differ in precisely this respect, we complemented the qualitative, hypothesis testing approach described so far with a quantitative, model fitting approach.

The models described in the Introduction were fit to the same categorization or similarity ratings, involved in the previous analyses. However, instead of using a derivative measure of performance as predicted variable, the models were fit to the raw categorization and similarity ratings obtained in the experiments. The predictor variable was derived from the importance ratings obtained in session 1 of the Experiments for each the properties entering in the descriptions. The predictor variable was therefore adapted to each participant, category and description used. In some probabilistic models, the importance ratings collected were simply added or subtracted depending on whether the corresponding property was asserted or negated in the description. In other models, the weights were then averaged over the number of properties in the descriptions. We also attempted to fit the results, using multiplicative rules. Such single-process models were compared with dual-process models involving special treatment of the necessary and/or defining properties. This special treatment usually consisted in attributing maximum weight to descriptions containing such information, irrespective of the number of properties in the descriptions and of the presence or absence of conflicting information.

One of the best accounts of the categorization and similarity judgments was obtained by assuming that, in all but one circumstance, participants simply averaged the weights of the properties in the descriptions. The one exception concerned definitions, to which the model assigned maximum weight in the categorization task, but not in the similarity judgment task. This model accounted for close to half of the variance over all experiments. This is not negligible considering that the model was fit to over 6000 data points, involving both subject and item variability, using only two estimated parameters: the intercept of the regression of obtained over predicted (categorization or similarity) ratings was 4.17 and the slope, 0.36. Although this model did better than many others, some of the models tested did about as well. So, on the basis of available evidence, it appears premature to commit oneself to a specific integration rule, be it the summation rule proposed by McNamara and Sternberg (1983), an averaging rule such as that proposed by Hampton (1988) or a multiplicative scheme (Hampton, 1995).

Another issue that remains open is whether some types of categories are represented and processed differently from others. There was some indication in the categorization data that definitions may not have been treated as differently from characteristic descriptions when they concerned artifact names than when they concerned well-defined and biological kind names. However, the lack of reliable category by description type interaction prevents rejection of the null hypothesis. One should be equally weary of asserting that all categories were treated alike since the statistical analyses performed on the regression coefficients were not very powerful, involving only 8 items per category type.

Finally, it must be stressed once again that the results presented here were obtained in conditions that gave the participants ample time to reach a decision. Such conditions make our interpretation of the results in terms of underlying theories quite plausible. By the same token, they impose caution about whether similar results would obtain in quick perceptual identification or in on-line language processing.

Footnotes

[1] We do not know any contemporary researcher who would argue that non-necessary or non-sufficient characteristic information will be ignored, especially when it is the only information available for categorization. However, this prediction is useful for distinguishing the classical view from dual views of categorization, described later.

[2] We chose the label “well-defined” because categories in the latter group were most likely to have a definition. However, there has been considerable debate about the existence, knowledge and use of definitions for all the category types involved in the present study.

[3] The purpose of the 0 rating was to exclude properties deemed ludicrous or tied to a different meaning of the category name presented (e.g., *triangle* as a geometric figure or as a love situation).

[4] These means were computed over participant-category pairs for which at least one necessary or defining property was identified. Being interested here in the size (and importance) of the sets of properties selected as necessary and especially as defining, the data of participants who did not select a single necessary or defining property were excluded from the results.

[5] Differences mentioned in the text are supported by statistical tests that were at least significant at the .05 level.

[6] The negative version of the properties were usually made by simply negating the main verb of the clauses. However, in order to prevent later decisions from being based solely on the surface form of the properties, antonyms were used whenever feasible (e.g., "is big" became "is small"; "is a man" became "is a women").

[7] The properties in the descriptions were sometimes composed of more than one proposition. If propositions are defined as the smallest units of meaning whose truth can

be verified, then one must consider the property “has along stem” (*tulip*), for instance, to be composed of two propositions: “has a stem” and “stem is long”. Measuring description size in terms of number of propositions instead of number of properties therefore provides a more precise measure of the amount of information in the descriptions. The results presented here are based on number of propositions as the measure of description size. However, analyses of categorization performance based on number of properties yielded results very similar to those presented.

[8] The same is true of models involving a multiplicative rule, such as found in exemplar models.

References

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

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

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 12-32.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory & Language, 34*, 686-708.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65*, 137-165.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*, 500-526.

McNamara, T. P., & Sternberg, R. J. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *22*, 449-474.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony, eds., *Similarity and analogical reasoning*. New York: Cambridge University Press.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge (Mass.): MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289-316.

Osherson, D., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9*, 35-58.

Pothos, E. M., & Hahn, U. (2000). So concepts aren't definitions, but do they have necessary or sufficient features? *British Journal of Psychology, 91*, 439-450.

Putnam. H. (1973). Meaning and reference. *Journal of Philosophy, 70*, 699-711.

Putnam, H. (1975). The meaning of meaning. In H. Putnam (Ed.), *Mind, language and reality, Volume 2*. London: Cambridge University Press.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony. (Eds.), *Similarity and analogical reasoning.* New York: Cambridge University Press.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122*, 468-486.

Saumier, D. (1993). *Are concepts all-or-none, or not?* Université de Montréal: Unpublished doctoral dissertation.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge: Havard University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81(3)*, 214-241.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22*, 377-386.

Table 1

Category names used in the Experiments (with English translation).

WELL-DEFINED TERMS		***BIOLOGICAL KIND***		***ARTEFACT TERMS***	
Célibataire	(Bachelor)	Carotte	(Carrot)	Avion	(Airplane)
Coup de circuit	(Home-run)	Chien	(Dog)	Chaise	(Chair)
Deux	(Two)	Érable	(Maple tree)	Maison	(House)
Eau	(Water)	Hirondelle	(Sparrow)	Marteau	(Hammer)
Grand-mère	(Grandmother)	Mouche	(Fly)	Ordinateur	(Computer)
Lundi	(Monday)	Pomme	(Apple)	Pantalon	(Pants)
Majorité	(Majority)	Truite	(Trout)	Revolver	(Revolver)
Triangle	(Triangle)	Tulipe	(Tulip)	Violon	(Violin)

Table 2

Percentage (%) of participants having identified necessary or defining properties.
Average number (N) and importance (Imp.) of the properties deemed necessary or defining for the well-defined, biological kind, and artefact terms over all experiments in the study.

Category type	Necessary			Defining		
	%	N	Imp.	%	N	Imp.
Well-defined						
Exp. 1	98	8.5	7.07	92	1.8	7.49
Exp. 2	97	8.4	7.44	95	1.9	7.92
Exp. 3	98	7.2	7.49	97	1.8	8.03
ALL	98	8.0	7.33	95	1.8	7.81
Biological kind						
Exp. 1	100	13.0	7.02	81	2.6	7.52
Exp. 2	99	13.0	7.02	93	2.9	7.71
Exp. 3	98	10.2	7.42	85	2.7	7.94
ALL	99	12.0	7.15	86	2.7	7.72
Artefact						
Exp. 1	99	11.2	7.03	89	2.9	7.58
Exp. 2	100	11.4	7.14	95	2.9	7.93
Exp. 3	100	9.4	7.55	93	2.7	7.85
ALL	100	10.6	7.24	92	2.8	7.78

Figure captions

Figure 1: Predictions of the classical view for descriptions made of **X** defining (**d**), necessary (**n**) or characteristic (**c**) properties that are either asserted (**+**) or negated (**-**).

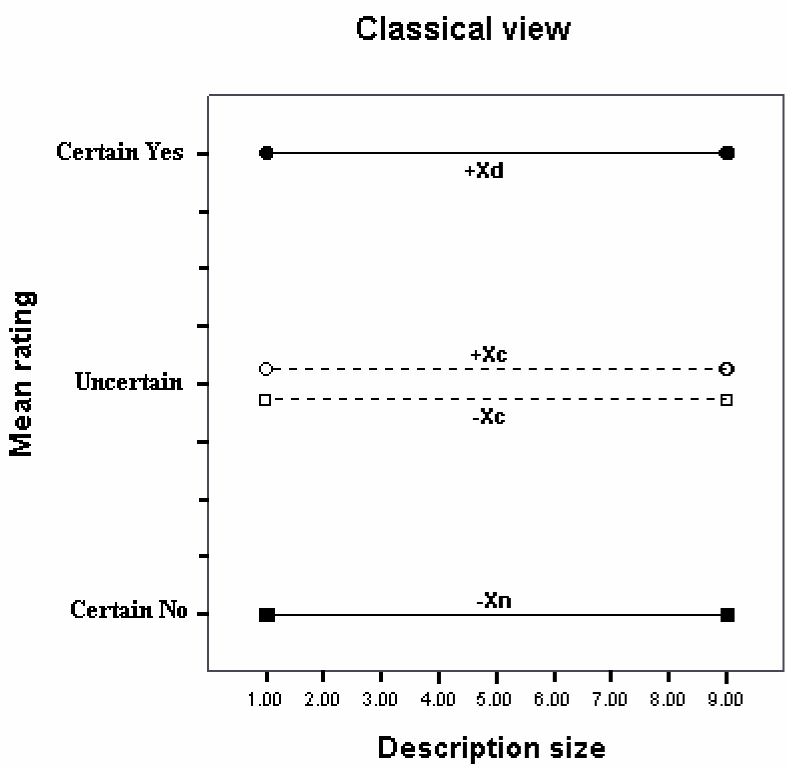
Figure 2: Predictions of a probabilistic view for descriptions made of **X** defining (**d**), necessary (**n**) or characteristic (**c**) properties that are either asserted (**+**) or negated (**-**).

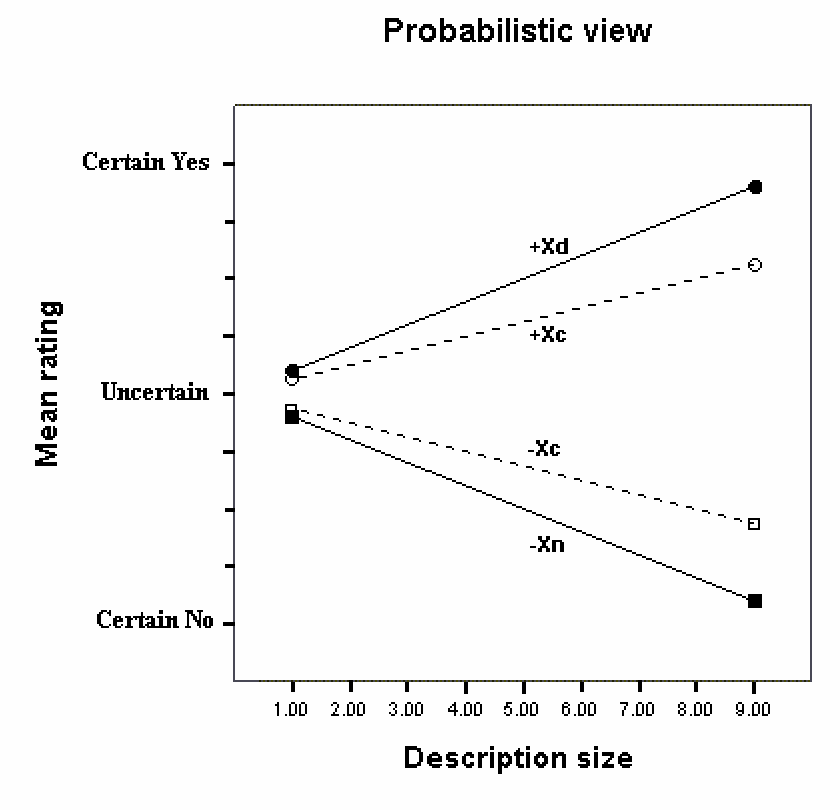
Figure 3: Predictions of a dual view for descriptions made of **X** defining (**d**), necessary (**n**) or characteristic (**c**) properties that are either asserted (**+**) or negated (**-**).

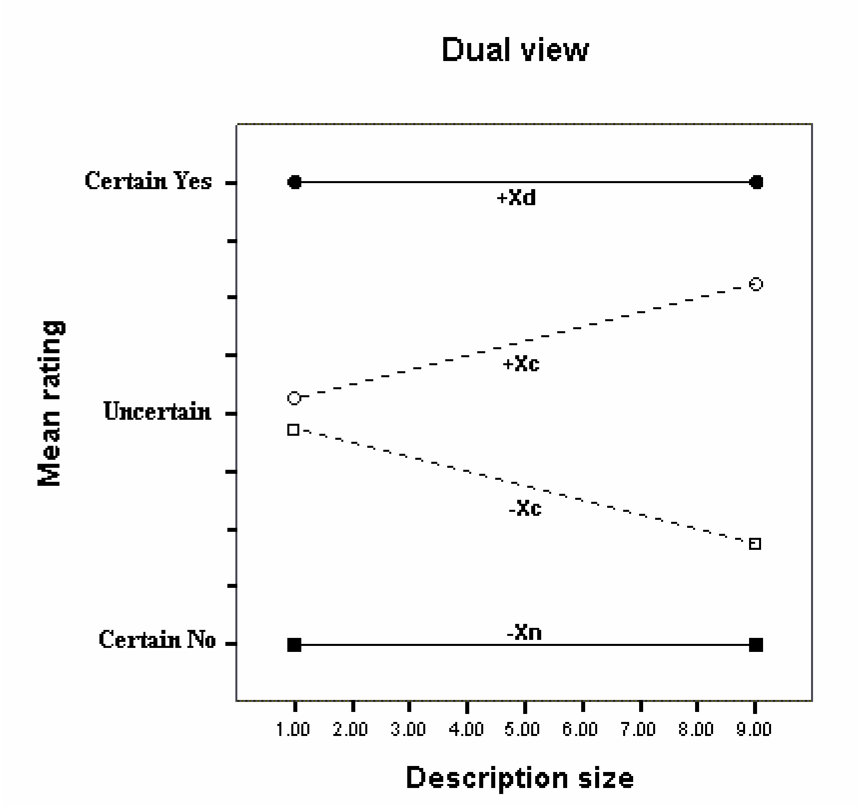
Figure 4: Mean categorization ratings, along with the mean intercepts and slopes obtained with the various description and category types (Experiment 1).

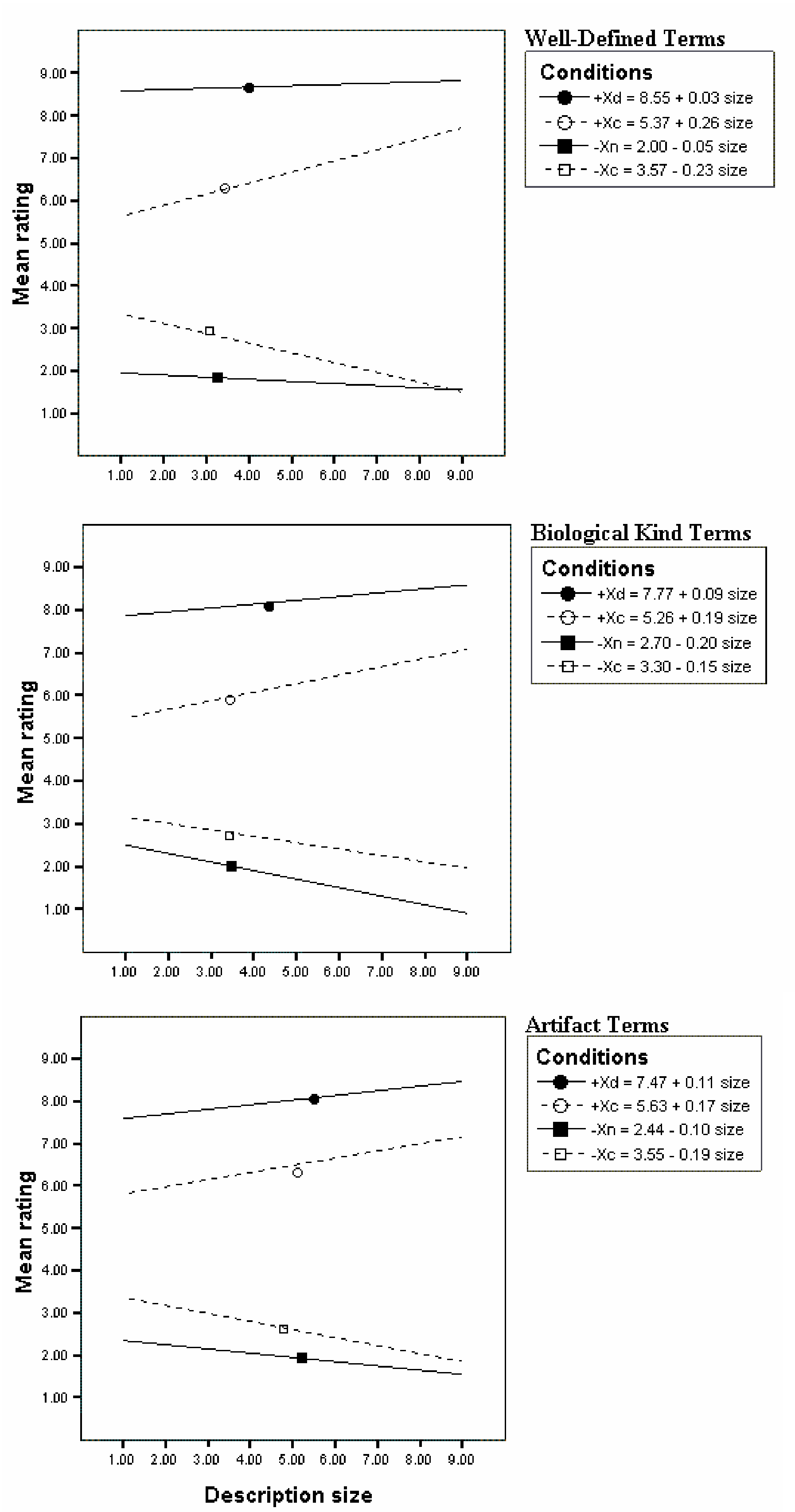
Figure 5: Mean categorization ratings, along with the mean intercepts and slopes obtained with the various description and category types (Experiment 2).

Figure 6: Mean similarity ratings, along with the mean intercepts and slopes obtained with the various description and category types (Experiment 3).

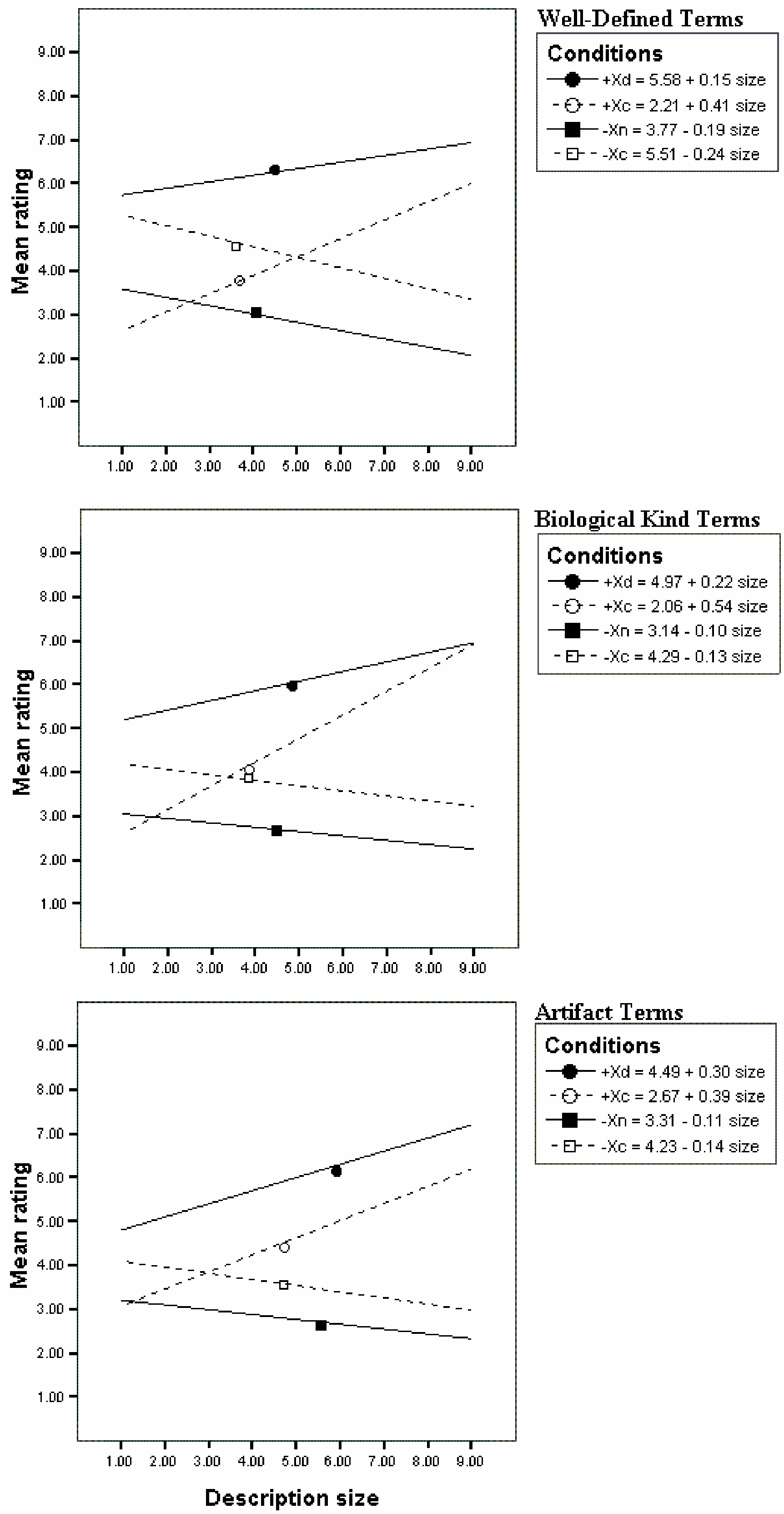

Classical view
Certain Yes
Uncertain
Certain No
Mean rating
+Xd
+Xc
-Xc
-Xn
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
Description size

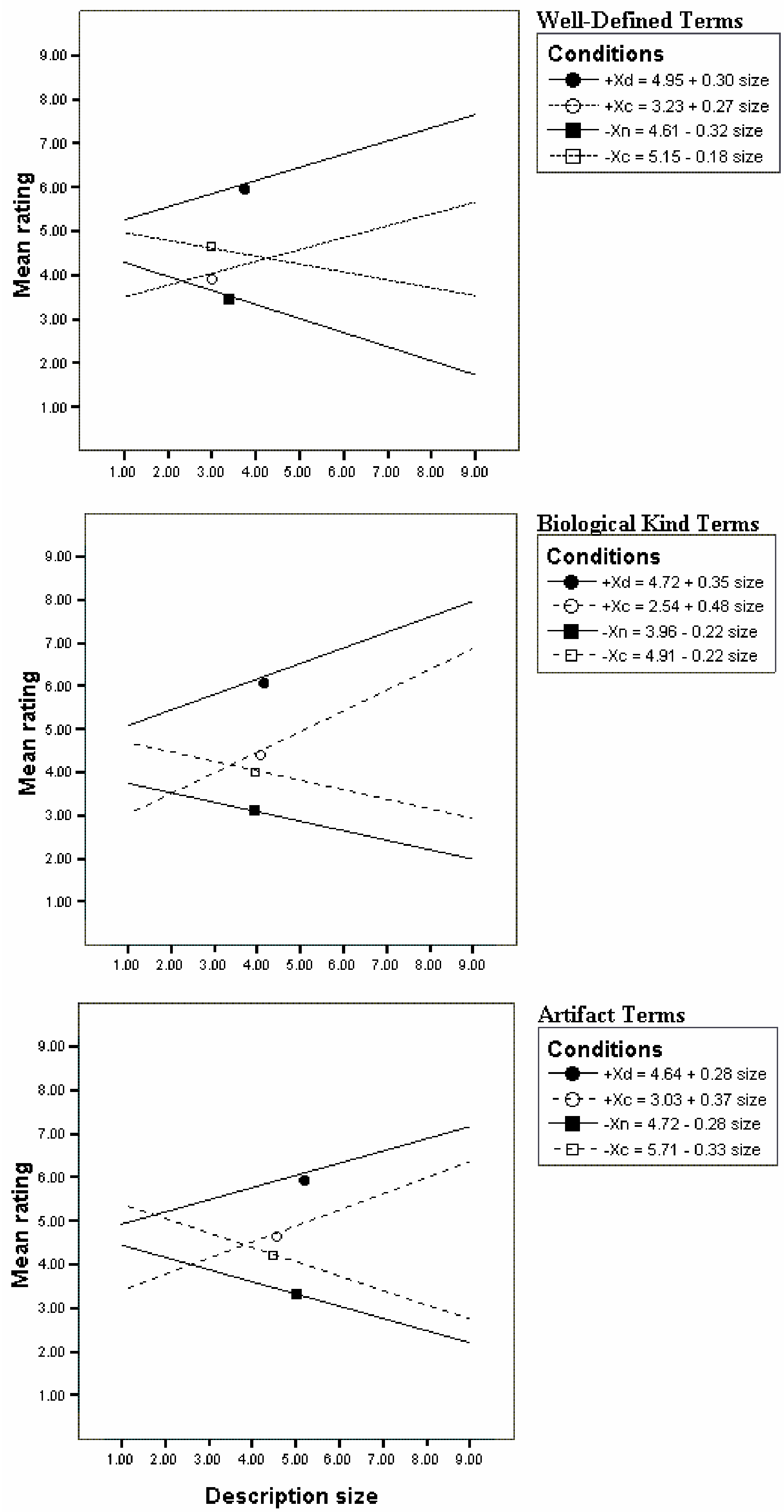

Probabilistic view
Certain Yes
Uncertain
Certain No
Mean rating
+Xd
+Xc
-Xc
-Xn
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
Description size


Dual view
Certain Yes
Uncertain
Certain No
Mean rating
+Xd
+Xc
-Xc
-Xn
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
Description size



Well-Defined Terms
Conditions
+Xd = 8.55 + 0.03 size
+Xc = 5.37 + 0.26 size
-Xn = 2.00 - 0.05 size
-Xc = 3.57 - 0.23 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Biological Kind Terms
Conditions
+Xd = 7.77 + 0.09 size
+Xc = 5.26 + 0.19 size
-Xn = 2.70 - 0.20 size
-Xc = 3.30 - 0.15 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Artifact Terms
Conditions
+Xd = 7.47 + 0.11 size
+Xc = 5.63 + 0.17 size
-Xn = 2.44 - 0.10 size
-Xc = 3.55 - 0.19 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Description size


Well-Defined Terms
Conditions
+Xd = 5.58 + 0.15 size
+Xc = 2.21 + 0.41 size
-Xn = 3.77 - 0.19 size
-Xc = 5.51 - 0.24 size
Mean rating
Biological Kind Terms
Conditions
+Xd = 4.97 + 0.22 size
+Xc = 2.06 + 0.54 size
-Xn = 3.14 - 0.10 size
-Xc = 4.29 - 0.13 size
Mean rating
Artifact Terms
Conditions
+Xd = 4.49 + 0.30 size
+Xc = 2.67 + 0.39 size
-Xn = 3.31 - 0.11 size
-Xc = 4.23 - 0.14 size
Mean rating
Description size


Well-Defined Terms
Conditions
+Xd = 4.95 + 0.30 size
+Xc = 3.23 + 0.27 size
-Xn = 4.61 - 0.32 size
-Xc = 5.15 - 0.18 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Biological Kind Terms
Conditions
+Xd = 4.72 + 0.35 size
+Xc = 2.54 + 0.48 size
-Xn = 3.96 - 0.22 size
-Xc = 4.91 - 0.22 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Artifact Terms
Conditions
+Xd = 4.64 + 0.28 size
+Xc = 3.03 + 0.37 size
-Xn = 4.72 - 0.28 size
-Xc = 5.71 - 0.33 size
Mean rating
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00
Description size