Memory & Cognition
2002, 30 (3), 340-352

Essentialist to some degree: Beliefs about the structure of natural kind categories

CHARLES W. KALISH
University of Wisconsin, Madison, Wisconsin

Previous research has provided conflicting evidence regarding the hypothesis that people are essentialists. Much of the evidence in favor of essentialism is based on demonstrating that categories are thought to have absolute membership. Although the hypothesis is often framed as an absolute claim about all categories of a certain type (e.g., natural kinds), it has generally been tested by making relative comparisons with a select sample. The present study assesses judgments of absolute structure across a range of categories. A further condition for essentialism is that the criteria for category identity be seen as objective rather than conventional. The results of three experiments based on these considerations do not provide support for essentialist claims. Few categories were judged to have essentialist structure, in terms of either absolute membership or objective criteria. Results are discussed in light of an alternative to the essentialist hypothesis that emphasizes a pragmatic view of categories.

There has been considerable interest in the proposition that people are essentialists about concepts and categories (Braisby, Franks, & Hampton, 1996; Coley & Luhmann, 2000; Diesendruck & Gelman, 1999; Gelman, Coley, & Gottfried, 1994; Hampton, 1998; Kalish, 1995, 1998; Keil, 1989; Malt, 1990; Medin & Ortony, 1989; Strevens, 2000). Psychological essentialism (Medin & Ortony, 1989) is the hypothesis that people represent categories as containing a core set of essential features that are both definitive and causally responsible for other, more peripheral properties. The literature describes empirical results both supportive of and conflicting with essentialist claims. Much of the debate has centered on people's conceptions of categories of living things (e.g., taxonomic concepts, such as BIRD or TIGER), often described as *natural kinds*. Without presuming to resolve all the issues, the purpose of this paper is to ask whether people are essentialists about natural kinds. Two facets of this question are addressed below. First, what kinds of empirical results and methods distinguish concepts represented as having essences (essentialized concepts) from those thought to have some other structure? Second, which categories have, or are supposed to have, essentialist structure?

This research was supported by a grant from NICHD (HD 37520) Thanks to Chris Lawson, Carrie Pritchard, Sean Shiverick, and Brian Viola for their assistance with data collection. Thanks also to Greg Murphy for helpful comments on an earlier draft. Correspondence concerning this article should be addressed to C. W. Kalish, Department of Educational Psychology, University of Wisconsin, Madison, 1025 W. Johnson St., Madison, WI 53706 (e-mail: cwkalish@facstaff.wisc.edu).

—Accepted by previous editorial team

Nominal and Real Essences

To address the question of results and methods, it is necessary to have a clear statement of the essentialist hypothesis. Essences are best understood by way of contrast with definitions. Definitions and essences both constitute categories; both essences and definitions are necessary and sufficient for category membership. However, unlike definitions, essences need not be explicitly represented. People may operate with an "essence placeholder" (Medin & Ortony, 1989) or a simple belief that there is some unknown essence to a category. Thus, people may believe that TIGER has an essence but have no commitment as to what that essence consists of. A second difference is that definitions are often arbitrary; one can artificially create a category by stipulating a definition. In contrast, essences are matters of objective fact. We discover which entities share an essence. Thus, essentialized categories can be evaluated as correct or incorrect. Scientists might discover that lions and tigers have the same essence. Our distinction between the two would then be revealed to be an error (see Braisby et al., 1996, for a discussion). In contrast, there are no facts about the world that could invalidate the consensually agreed upon notion of BACHELOR. To highlight similarities and differences, definitions are often described as *nominal essences*, in contrast to truly existing *real essences* (Locke, 1707/1961). In testing hypotheses about essences, it is critical that methods distinguish between nominal and real essences.

Research on essentialism has explored whether membership in categories is all-or-none or a matter of degree (Coley & Luhmann, 2000; Diesendruck & Gelman, 1999; Hampton, 1998; Kalish, 1995). Such a distinction is a primary source of evidence for claims that people are essentialists with respect to taxonomic concepts (e.g., BIRD, TIGER; see

Diesendruck & Gelman, 1999). Particularly compelling is evidence that even if people are not themselves able to determine category identity with certainty, they generally believe that an expert can do so for taxonomic concepts, but do not believe such absolute criteria exist for concepts of artifacts (Coley & Luhmann, 2000; Malt, 1990). The explanation for this result is that people expect experts (e.g., scientists) to be in possession of information about essential properties. Because artifacts are not thought to have essences, experts are in no privileged position to determine category membership. However, such findings do not establish that people believe in real essences.

Both nominal and real essences may provide absolute standards for category membership. Moreover, the quality of deferring to experts for the final word on category boundaries is not restricted to kinds with real essences; division of linguistic labor holds quite generally (Burge, 1979). In addition to the intuition that there are clear and absolute criteria for membership, a commitment to real essences further entails the belief that these criteria are objective rather than conventional. Kalish (1998) found that people believed there was a single correct way to classify animals but were more accepting of diverse schemes for categorizing artifacts. This pattern is consistent with a belief that natural kinds have real essences. However, only a few categories were examined in that study, and the results revealed that some artifact classifications were treated as objective. Thus, it remains something of an open question as to whether people believe that categories have real essences or merely hold that some categories have absolute membership.

Natural Kinds

Given the above characterization of essentialism, which categories might be thought to have real essences? A frequent suggestion is that natural kinds have real essences, whereas artifacts do not (Ahn, 1998; Barton & Komatsu, 1989; Diesendruck & Gelman, 1999; Kalish, 1998; Keil, 1989). In philosophic usage, *natural kind* just means *category with a real essence* (see Kripke, 1972). If there is an empirical question whether natural kinds have real essences, psychologists must be operating with a different notion. Unfortunately there is no agreed upon characterization (see Kalish, in press, for a discussion). On the one hand, there is a relatively broad definition—roughly, categories of naturally occurring objects (Gelman, 1988; Smith, 1995). However, naturally occurring objects may be categorized in a near infinite number of ways, making it unlikely that all categories would share a common structure. A reasonable condition to add might be that natural kinds are those categories of naturally occurring objects that are not clearly arbitrary or based on social/instrumental criteria (e.g., defined by human interaction or usage, such as PET). This is the characterization of natural kind used in the present study. An alternative proposal is that natural kinds are those that figure in the causal laws of theories (Gelman et al., 1994; Keil, 1989). The class of categories predicted to have essential structure is limited to those that are *theory-laden* (Gelman et al., 1994). On this view, for example, taxonomic categories of living things are said to be natural kinds because they are central to naive theories. A yet more specific construal is that essentialism holds only for taxonomic categories, not for other theoretically important concepts. For example, Atran (1987) argues that taxonomic categories of living things are subserved by a dedicated cognitive module that produces essentialist structure.

Combining the different characterizations of natural kinds, it is possible to develop a set of predictions. The hypothesis that categories have essences applies by consensus to taxonomic categories of living things; if any categories have essential structure, they should. A second level of the prediction is that other theory-laden categories will also be essentialized. Thus, kinds that figure in naive chemical, physical, and, perhaps, geological and ecological theories may be thought to have essential structure. Finally, the broadest prediction is that most nonarbitrary noninstrumental categories of naturally occurring objects will be essentialized. To begin to explore these hypotheses, it is necessary to test a range of categories. Some construals predict differences in structure among categories of natural objects; some predict homogeneity.

In addressing the question of category structure, the usual strategy has been a comparative approach. Researchers compare the structures of natural kind and artifact categories. As a procedure for addressing absolute questions about category structure, this strategy faces two limitations. The first is that essentialism could be more characteristic of one kind of category than of another but still really hold true of neither. For example, membership in natural kind categories may be judged to be relatively more absolute than membership in artifact categories but still be judged to display a significant amount of gradedness. Some absolute standards for category structure are necessary in addition to the comparative measures (see Kalish, 1995, 1998).

A second difficulty with the ways the questions have been framed is that category structure applies, ultimately, to individual categories, not to groups or types of categories. Thus, to say that natural kinds have essences but artifacts do not is ambiguous. Is it that all natural kinds have essences but no artifacts do? Or is it that, on average, the level or frequency of essentialism is higher for one set than for the other? If it is the latter, problems of sample selection become crucially important. This issue was recently raised by Diesendruck and Gelman (1999), who criticize Kalish (1995) for drawing conclusions based on a restricted or odd sample of natural kinds. In the literature, there has been very little discussion of sampling; indeed, there is little discussion of what the population is from which the sample is to be taken. One exception is the recognition that measures of category structure should be made conditional on typicality (see Hampton, 1998). It makes little sense, for example, to compare highly typical natural kinds with very atypical artifacts. Any model would predict higher levels of absolute categorization for natural

kinds in this case. Yet, is typicality the only covariate to consider? Are the kinds of categories otherwise homogeneous?

This review of the literature on essentialism has highlighted several limitations of existing research. Three issues must be addressed in order to provide a clear test of the hypothesis that people are essentialists about natural kind categories. First is the question of whether all natural kind categories have the same structure, or whether essentialism is more characteristic of some than of others (e.g., taxonomic vs. others). The experiments described in this paper address this question by exploring a wide range of natural kind categories. A second question is whether absolute levels of essentialism are high, or whether some categories appear essentialized only in comparison. The experiments below address this question by including comparison categories known to have or lack essential structure. The third question is a measurement one. Do measures of essentialism differentiate between real and nominal essences? Demonstrating that some categories have absolute membership criteria is not sufficient to establish essentialism; it is also necessary to show that the criteria are objective, rather than conventional. The experiments below address this question by testing both absolute membership and intuitions about objectivity.

EXPERIMENT 1

To count as having an essence, a category must have two properties. First, essentialized categories have absolute rather than graded membership. Second, essentialized categories are objectively rather than conventionally determined. In Experiment 1, both aspects of category structure were examined. Experiment 1 also addressed the sampling question: How common are categories with essences? The strategy adopted in Experiment 1 was to derive a sample of potentially essentialized categories from an independent source—in this case, a corpus of word frequencies. The most frequent terms for natural kinds were selected as stimuli (see below for details on selection). Although word frequency is only one possible way to select a set of categories, it does have prima facie validity as a strategy for achieving a representative sample. This procedure is also relatively unbiased. Rather than generating a list based on some assumptions about the nature of essentialized kinds, selecting from a list of word frequencies assures that a range of kinds will be considered. However, because of their centrality to claims about essences, taxonomic categories were intentionally oversampled for this experiment (Gelman et al., 1994; Keil, 1989). Finally, to address the question of absolute levels of essentialist beliefs, the experiment also included comparison categories.

Method

Participants. Nineteen students at a large Midwestern university participated in the experiment. The students received course credit for participation.

Design. Stimuli for the experiment were drawn from a list of word frequencies compiled from the British National Corpus (Kilgarriff, 1998). The full list contained 6,318 words. The 40 most frequent terms for natural kinds were selected to serve as stimuli. Terms were included if they designated naturally occurring objects or substances. Excluded from the list were words referring to body parts (e.g., *hand*, *head*), instrumental categories (e.g., *pet*), and labels for groups of people (e.g., *man*, *girl*) because they also designate social categories.[1] It was judged that inclusion of these types of items would bias the list against essentialist interpretation; dropping them was a conservative move to present the best chance for finding evidence of essentialism. The selection of natural kind terms was divided into two phases. In the first phase, the 20 most frequent terms denoting taxonomic categories were selected. The second phase involved selection of the 20 most frequent terms denoting other types of natural kinds. Of the nontaxonomic natural kinds, 5 designated biological entities, 15 were nonbiological. In addition, 6 items were included to provide clear cases of graded (fuzzy) and all-or-none (well-defined) categories. A complete list of the items is presented in the "Category" column of Appendix A.[2]

Each item was presented as the subject of a debate between two experts. Items described two experts coming across an unfamiliar sample or individual. The sample was described with a superordinate label (e.g., *substance*). The experts disagreed whether the object was a member of a given category. The use of disagreeing experts served two purposes. First, it encouraged the participants to consider truly difficult cases (over which even experts might disagree). Second, identifying the relevant type of expert reduced any ambiguity over the interpretation of the category term. For example, it was two geologists who disagreed whether a substance was oil. The types of experts and the superordinate labels used for each category are indicated in Appendix A.

The participants were asked to respond to three statements. The first statement described the category as graded: "There is a continuum from pure ideal instances to imperfect partial instances. Some things may be truly intermediate between a BIRD and not." Across items, statements differed only in the category label used. The second statement described the category in question as absolute: "Even a BIRD that is strange or unusual is still 100% a BIRD. People may be confused or have a hard time telling, but there are no partial BIRDS." In both cases, the participants rated their agreement with the two statements on a 20-point scale. Finally, a third question assessed whether the participants judged the distinction between category members and nonmembers to be a matter of discovered fact or a matter of decided-upon convention. The actual question was the following. "Is the exact distinction between imperfect BIRDS and non-BIRDS a convention or a matter of fact? Is the boundary something that experts DECIDE or is it something they DISCOVER?" The participants rated whether the distinction was "a discovery of fact" or a "decision or convention" on a 20-point scale.

Procedure. The participants were tested in small groups in a classroom equipped with 12 desktop computers. The participants were informed that they would be evaluating categories as absolute or graded. The two alternatives were illustrated using EVEN NUMBER and HEAVY as examples. Following these instructions, the participants were presented with the 46 items in random order. Each item was presented as a separate "screen," with the three statements appearing in turn. The order of rating the statements was always, first, the graded statement, then the absolute statement, and then the question about conventions. The participants made ratings by moving a marker toward one of the two labeled ends of a continuous region (the response measure looked and functioned like a standard scroll-bar).

Results and Discussion

By reverse coding the gradedness and convention measures, high scores on each represent a greater commitment to essentialism. As was discussed above, categories were grouped into five types. The two control types were fuzzy

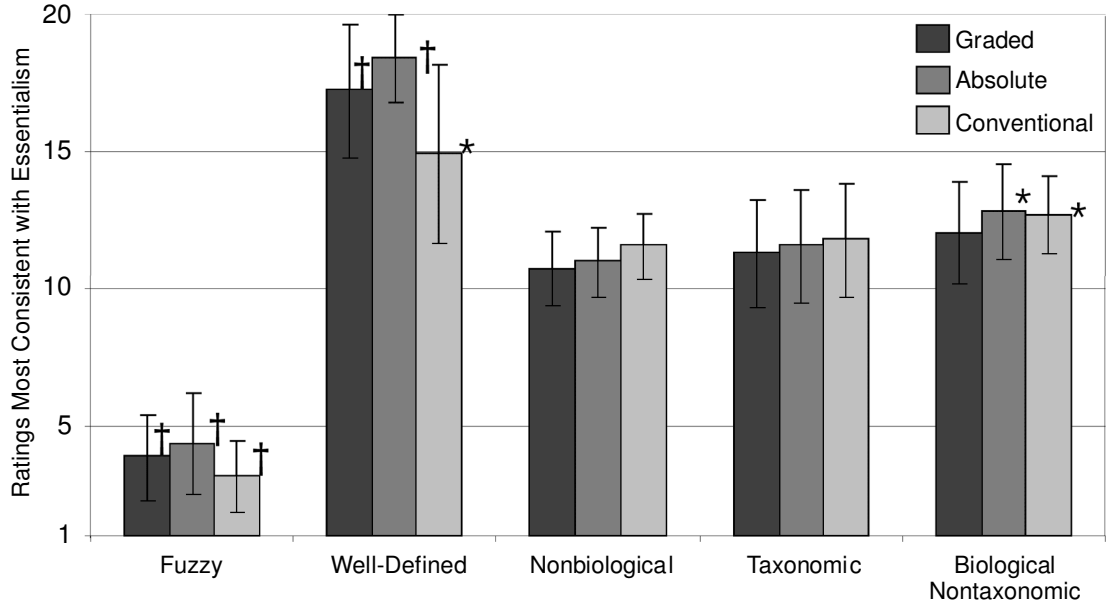

Figure 1. Mean ratings of graded and absolute category structure and conventionality in Experiment 1. Higher scores indicate responses more consistent with essentialism. All measures ranged from 1 to 20. *Different from chance (10.5) at $p < .05$. †Different from chance at $p < .01$. Error bars represent 95% confidence intervals.

and well defined (with the unexpected "logically valid" item excluded from analyses; see note 2). Natural kinds were divided into three types: taxonomic, biological nontaxonomic, and nonbiological. Figure 1 presents the mean scores on the three measures for each category type. (Mean scores on the three measures for each individual item are listed in Appendix A.) An analysis of variance (ANOVA), with measure and category type as within-subjects variables, revealed a main effect of category type [$F(4,72) = 51.1$, $MS_e = 25.6$, $p < .001$]. Pairwise comparisons showed that the fuzzy items were scored significantly lower than all other items and that the well-defined items were scored significantly higher (Tukey's HSD, $p < .01$). No other pairwise comparisons were significant.

As can be seen from Figure 1, the three natural kind groups were intermediate between the well-defined and the fuzzy groups. Moreover, the different groups of natural kind categories did not appear to be significantly different from each other. Ratings for the natural kinds were close to the midpoints of the scales (10.5). It was not the case, however, that the participants used only a restricted part of the scales. Fourteen of the 19 participants used almost the entire range of all three scales for natural kind categories (ranges at least as large as 3–18). There was significant variability across items and/or participants.

In general, there was a high correlation between items' mean scores on each of the three measures (Pearson's r for gradedness vs. absolute, .98; gradedness vs. convention, .86; absolute vs. convention, .84). The ANOVA described above did reveal an interaction between measure and category type [$F(8,144) = 2.6$, $MS_e = 5.4$, $p < .05$]. Scores on the convention measure for well-defined items were somewhat lower than scores on the other two, although the post hoc contrast did not reach statistical significance [$F(14,144) = 1.4$, $MS_e = 5.7$, n.s.; Scheffé's test]. This disassociation is explored more fully in Experiment 3.

The second set of analyses focused on individual differences. One question is whether individuals have different propensities to give essentialist responses for the natural kind items. Given the high correlations between measures (graded, absolute, and conventionality), *essentialism scores* were computed by taking their average. A significance test of individual patterns may be constructed by assuming that the midpoint of the scales (10.5) is the level expected by chance. A pattern of responses consistent with essentialism may be defined as responses equal to or above 10.5 for 26 or more of the 40 natural kinds, assuming $p(\geq 10.5 = .5)$, $p(26 \text{ or more out of } 40) < .05$). Eight of the 19 participants showed the essentialist pattern. Six participants consistently (15 or more out of 20) rated the taxonomic items at or above the midpoint. Using these weak criteria, fewer than half of the participants showed evidence of essentialism for natural kinds.

To explore item differences, 95% confidence intervals were computed for each item. Confidence intervals for the natural kind items are shown in Figure 2. Confidence intervals were rather large, reflecting the individual differences. There did not appear to be a sharp distinction between essentialized and nonessentialized categories: Confidence intervals increased smoothly. Only seven natural kinds (HUMAN, EGG, AIR, WATER, CELL, FIRE, OIL; no taxonomic categories) had confidence intervals with lower bounds greater than 10.5. Post hoc inspection of the confidence intervals suggested that there might be a category

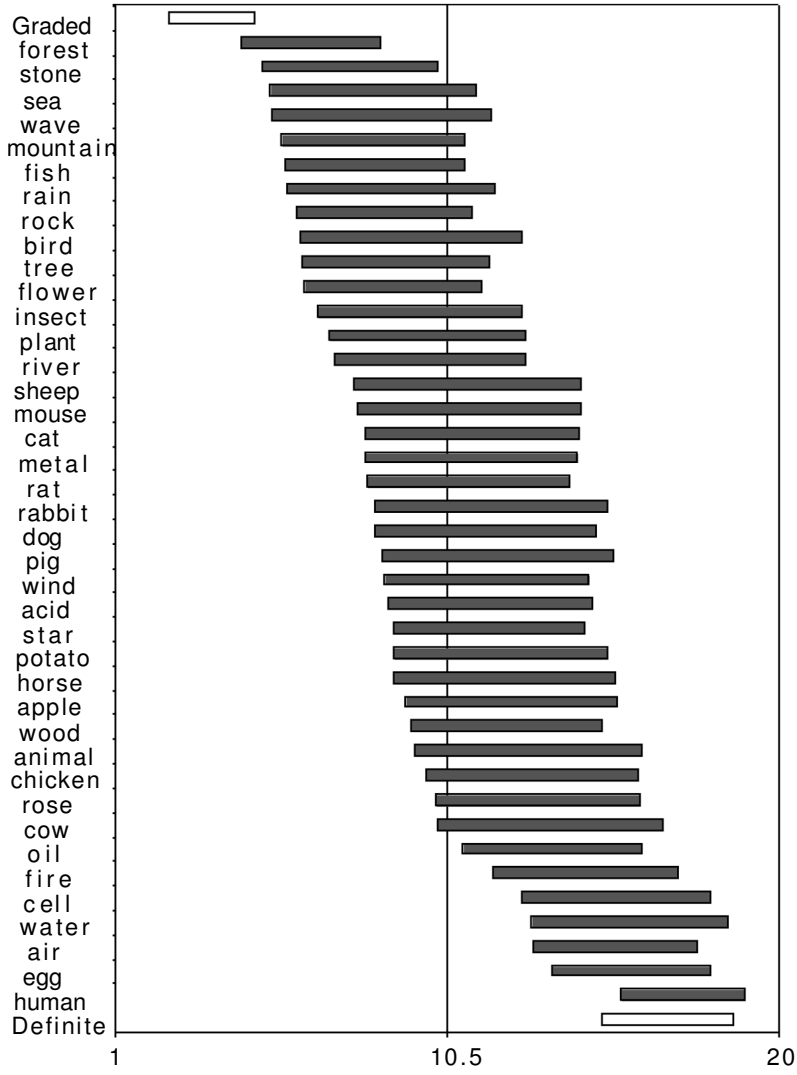

Figure 2. Ninety-five percent confidence intervals for essentialism scores from Experiment 1. Open bars represent the confidence intervals for the means for fuzzy and well-defined items.

level effect for taxonomic stimuli. Superordinate kinds received lower essentialism scores than did basic level kinds [$t(18) = 3.0$, $p < .01$, $M_{basic} = 11.7$, $M_{super} = 9.1$].

Overall, the participants assigned ratings to categories that indicated a moderate degree of essentialism. The intermediate status of natural kinds was demonstrated with respect to two sets of reference categories (well defined and fuzzy). Across three measures, natural kinds were rated as more essentialist than fuzzy categories, but less essentialist than well-defined categories. There was little evidence of subtypes within natural kinds. An examination of the absolute levels of ratings for natural kinds also suggests a moderate degree of essential structure. Again, across three measures, ratings for natural kinds were near the midpoints of the rating scales.

The results of the present experiment seem somewhat inconsistent with the results of prior studies. In particular, Malt (1990) and Coley and Luhmann (2000) both report high levels of essentialist responses for taxonomic categories, using dispute resolution paradigms. In the previous studies, participants have been asked whether an expert could resolve a questionable or ambiguous case. For taxonomic kinds, participants generally reported that an expert could make a definite determination of an object's category membership. In contrast, the participants in the present experiment did not expect experts to be able to resolve questionable instances. One possibility is that the design of Experiment 1 may have encouraged the participants to imagine fantastic or bizarre cases. Previous studies have asked whether an expert may resolve a dispute between

(or confusion among) nonexperts. In contrast, Experiment 1 involved disputes between experts. Any object that is challenging even to experts must be a truly unusual case. Just such unusual cases are the tests of essentialist claims: Essentialism implies there are no truly intermediate instances of categories. Nonetheless, the structure of the task may have encouraged the participants to think of fanciful examples that were not believed to be actually possible. To address this possibility, a second experiment was conducted, contrasting expert and nonexpert judgments.

EXPERIMENT 2

Experts know more about categories and are in positions of power to determine standards for correct classification. To begin to test the essentialist claim, it is necessary to decouple knowledge from authority. If an expert's role in determining classification is akin to stipulating a convention, such authority should be culturally specific. Just as people in other nations are not bound by our traffic laws, so to they would not be "bound" by our categorization practices. In contrast, if experts are references for determining category identity because of their superior knowledge of facts, those facts should be relevant across cultural differences. Expert knowledge of essences may reveal categorization practices to be wrong in other cultures, as well as in the expert's own.

In Experiment 2, participants were asked to make judgments about cultural differences in categorization practices. The goal was to distinguish expertise effects owing to superior knowledge from expertise effects owing to (culturally specific) authority. A second advantage of the method is that cultural differences in categorization practices are relatively unexceptional. It is not odd to think of some group of people categorizing bats with birds, whales with fish, or stars with planets. In contrast to the expert disagreements in Experiment 1, cultural differences are common enough without thinking about science fiction or fantastical scenarios.

Method

Participants. Thirty-three students at a large Midwestern university participated in the experiment. The students received course credit for participation. Only 1 participant had taken more than a single college-level biology class at the time of the experiment.

Design. The items included were the same as those used in Experiment 1. Six well-defined items were added to the set to provide a stronger comparison group with natural kinds. Four items involved individual identity—whether two fingerprints, hairs, and blood samples came from the same person, and whether two people had the same father. Individual identity is assumed to be essentialized: It is an absolute matter of objective fact whether some experience involves a single individual person or two different people. Proper names (denoting individuals) are often taken as the paradigmatic cases of terms with essentialist structure (Kripke, 1972). The fifth and sixth items were mathematical (whether equations equaled 8, and equaled each other), to replace the logical validity item of Experiment 1.

Each item was presented as a difference in categorization practice between a group of scientific experts and a people called the "Farnhi." The participants were told that "the Farnhi have not had much contact with Western science and have their own beliefs about the way the world works." In each instance, the objects of consideration were a pair of individuals (described with a superordinate term). One group judged that both individuals were the same kind of thing. The other group judged that one individual was a member of a category, whereas the other was not. An example of an item is the following:

> Our experts and some people from Farnhi are considering two animals. The experts and the Farnhi disagree about what kind of things they are. The Farnhi think both the animals are pigs. For them the two animals are the same kind of thing. Our experts think one of the animals is a pig and one isn't. For the experts, they are different kinds of animals.

Disagreements involved both the appropriate label to apply to the individuals and judgments about whether they were the same kind of thing or not. The participants were asked to make two ratings. The first asked for a judgment of whether there was a single correct way to resolve the dispute (objectivity). The participants made this rating on a 20-point scale, with endpoints labeled as follows: "One group is wrong. It is a matter of fact what is a pig and what's not" and "Different answers are acceptable depending on people's perspectives or values." A second judgment asked about gradedness, using a 20-point scale with endpoints labeled "Some animals may be sort of in between a pig and not" and "Something just is a pig or not, though it may be hard to tell. There is no 'in between.'"

Procedure. The procedure was the same as that used in Experiment 1. The order of presentation of the disputants (expert, Farnhi) and their sides of the dispute (both individuals the same vs. one different) were randomized across participants and items.

Results and Discussion

The results from Experiment 2 were generally consistent with those of Experiment 1. The participants gave ratings representing essentialist views for the well-defined items. Fuzzy items were not essentialized. All other items received intermediate scores, averaging in the middle of the scales. Mean ratings for the five types of items are presented in Figure 3.

The objectivity and gradedness measures were highly correlated ($r = .97$). Therefore, objectivity and (reverse-scored) gradedness ratings were averaged to produce essentialism scores. A within-subjects ANOVA revealed significant differences in essentialism scores by category type [$F(4,72) = 51.1$, $MS_e = 25.6$, $p < .001$]. Pairwise comparisons showed that the fuzzy items were scored significantly lower than all other items and that the well-defined items were scored significantly higher (Tukey's HSD, $p < .01$). Nontaxonomic biological items received higher scores than did nonbiological items. No other pairwise comparisons were significant.

There were substantial individual differences in patterns of essentialism scores. Few participants, though, reliably gave high scores to natural kinds. A pattern of rating 26 or more of the 40 natural kinds at 10.5 or higher differs from chance, with $p < .05$. Ten participants showed this pattern (the number is unchanged when considering only 15 or more out of the 20 taxonomic categories). Twenty-four participants consistently rated well-defined items above the midpoint (7 or more out of 8). There was also variability across individual items. Confidence intervals (95%) had the same graded pattern as that shown in Figure 2.

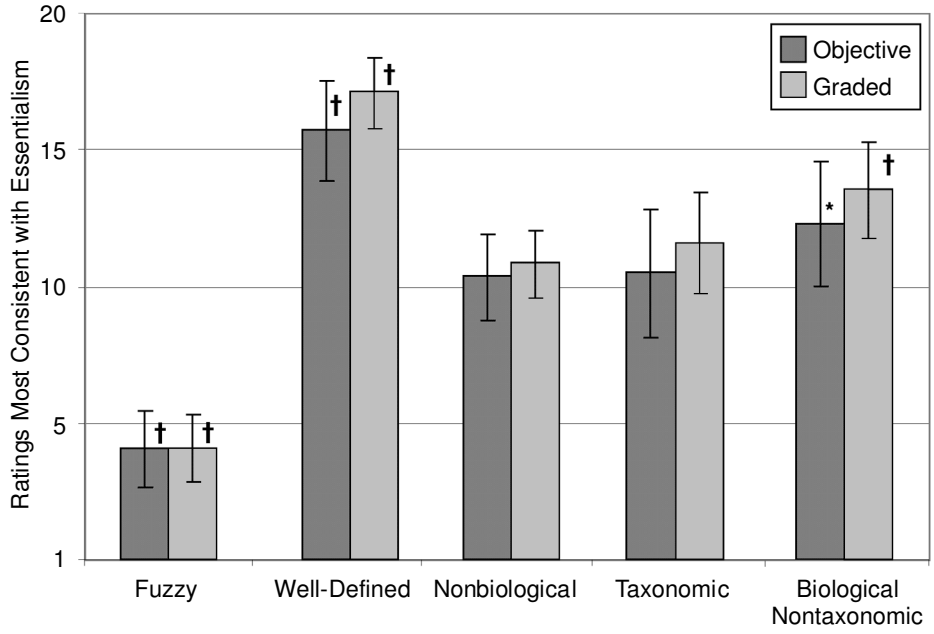

Figure 3. Mean ratings of category structure and objectivity in Experiment 2. Higher scores indicate responses more consistent with essentialism. All measures ranged from 1 to 20. *Different from chance (10.5) at $p < .05$. †Different from chance at $p < .01$. Error bars represent 95% confidence intervals.

Only eight natural kinds had lower bounds greater than 10.5 (CELL, WATER, EGG, ACID, HUMAN, FIRE, METAL, SHEEP). As compared with Experiment 1, the absolute levels of essentialism were somewhat higher, although the patterns of ratings across items were generally consistent. In particular, taxonomic categories were underrepresented among the natural kinds with high essentialism scores.

The results of Experiment 2 provide converging evidence for the conclusions from Experiment 1. Most people did not have consistently strong essentialist intuitions about taxonomic categories of living things. People did make judgments consistent with essentialism—in this experiment, about matters of mathematical fact and individual identity. The participants did not extend essentialist judgments to natural kinds. The reluctance to essentialize does not seem to be attributable to task demands that would encourage consideration of outlandish or science fiction examples. The procedure in Experiment 2 involved the relatively commonplace phenomena of cultural differences in categorization practices. Moreover, had participants been answering in terms of fantastic possibilities, such reasoning should have affected their judgments about well-defined items, as well as about categories of natural kinds.

EXPERIMENT 3

The results of Experiments 1 and 2 seem to contrast with past research contrasting essentialist beliefs about natural kind with the lack of such beliefs about artifact categories. Perhaps it is only relative to artifacts that natural kinds are essentialized. A somewhat more worrisome possibility is that the inclusion of artifact categories as a within-subjects manipulation affects ratings of natural kinds. The inclusion of both kinds of categories may induce a contrast response in which people accentuate the differences between stimuli. Experiment 3 explored whether evidence for essentialist views of natural kinds might be found when artifact categories are available as a contrast case.

Experiment 3 focused on whether categorization was seen as a matter of convention or a matter of fact. Among the measures that have been used to distinguish matters of convention from matters of fact, two are particularly appropriate for present purposes (Turiel, 1983; see Kalish, 1998, for application to categorization). A first distinction is between those things that are or are not legitimately alterable by authorized decision. We may ask whether assignment of an individual to a category is (legitimately) a matter of decision: May the properly authorized experts just decide how something should be categorized? A second measure concerns the legitimacy of alternative practices, akin to that used in Experiment 2. Is diversity in category membership judgments (e.g., across linguistic or cultural communities) acceptable or not? Experiment 3 used these two measure to explore intuitions about the nature of experts' judgments for natural kinds and artifacts.

Method

Participants. Eighteen students at a large Midwestern university participated in the experiment. The students received course credit for participation.

Design. The participants heard about experts confronting objects appearing intermediate between two categories. An example of an item is the following: "A committee of expert biologists was considering a plant that seemed to be sort of halfway between a marigold and a dandelion." The participants were then told that the group had been empowered to determine the identification of the plant and were asked, "How should the committee go about reaching a decision?" The measures for the experiment were ratings of the appropriateness of two ways of proceeding. The first option, headed "Investigate," was described as follows. "The experts could do some more scientific studies to discover whether the plant shares the essential biological properties of marigolds or dandelions. The way to resolve the dispute is to discover more facts about the plant's biology." The second option, headed "Legislate," was described as follows. "The problem is to establish a standard or rule. The experts have to choose whether it is more useful to call the plant a marigold or a dandelion. Which way they decide isn't so important as having a consistent policy." After rating the two alternatives, the participants were then asked a question probing their intuitions about diversity in decisions about the ambiguous item. This item was presented as follows:

> Now imagine that another group of experts has come to the OPPOSITE answer of this first group. One group of experts says the plant is a marigold, the other that it is a dandelion. Must one group actually be wrong? Is one way of answering the question right and the other wrong (even if we may not know which is which)? Or could the facts ultimately be consistent with both answers? Is this a matter of interpretation or preference?

The participants answered by indicating on a scale of 1–20 whether "one group is wrong, there is one right answer" or "either answer may be acceptable. It is a matter of perspective or opinion." Order of judgments was maintained across items: The investigate and the legislate options were presented simultaneously, followed by the diversity question.

Items. The items and characterizations were drawn (with some adaptation) from Malt's (1990) study of deference to experts. Malt had empirically established pairs of categories that were viewed as quite similar but nonetheless distinct. The eight natural kind items in the present experiment were the same eight pairs as those identified by Malt. Malt describes seven artifact pairs, of which one (*hide-and-seek–tag*) was judged to be a nominal kind for the purposes of this experiment. Three more artifact items were developed. These items were examples of scientific instruments: *thermometer–barometer*, *cyclotron–laser*, and *magnetic resonance imaging (MRI) scanner–positron emission topography (PET) scanner*. Inclusion of instruments allowed the questions and committees to be scientific ones and, thus, comparable to the natural kinds. The nominal kinds were five drawn from Malt. To this list was added the hide-and-seek item (originally designated an artifact) and a pair involving *marriage–partnership*. Note that the nominal kinds included two mathematical categories. The items included in the experiment are presented in the "Category" column of Appendix B.

Procedure. The participants were tested in small groups in a classroom equipped with 12 desktop computers. The instructions described the task as one of evaluating how experts should come to a decision about categorizing unfamiliar objects. The distinction between discoveries of fact and stipulations of convention was illustrated by the examples of determining whether the sun rotates around the earth and determining whether people should drive on the left side of the road. Following these instructions, items were presented one at a time in randomized order.

Results and Discussion

Responses were coded so that higher scores indicate ratings more consistent with essentialism—agreement with the investigate option, disagreement with the legislate option, and rejection of multiple alternatives for the diversity question. Figure 4 presents the mean ratings for each of these three measures. Mean ratings were subjected to an ANOVA, with category type (natural kind, artifact, nominal kind) and measure (investigate, legislate, diversity) as within-subjects variables. Both main effects were signifi-

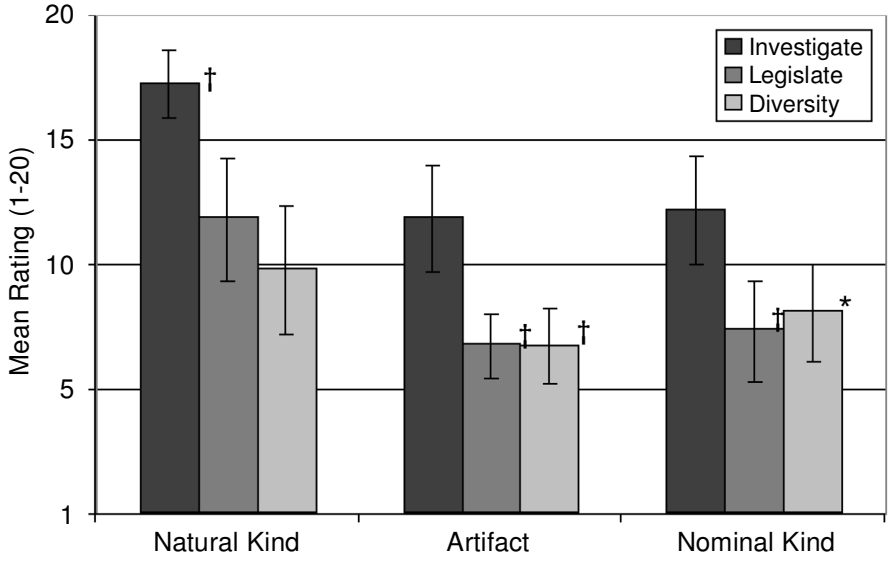

Figure 4. Results of Experiment 3. Evaluations of investigate and legislate options for resolving disputes about category membership. Also plotted are ratings indicating rejection of diversity in criteria for category membership. Higher scores indicate responses more consistent with essentialism. All measures ranged from 1 to 20. *Different from chance (10.5) at $p < .05$. †Different from chance at $p < .01$.

cant, as was the interaction [measure, $F(2,34) = 22.8$, $MS_e = 22.5$, $p < .001$; category type, $F(2,34) = 14.8$, $MS_e = 21.2$, $p < .001$; interaction, $F(4,68) = 4.1$, $MS_e = 3.8$, $p < .01$]. Across category types, investigate scores tended to be higher than others (Tukey's HSD, $p < .01$). Across measures, scores for natural kinds were higher than scores for artifacts and nominal kinds (which did not differ; Tukey's HSD, $p < .01$). The interaction derived from a smaller category type difference for the diversity measure than for the other two (see Figure 4). Of particular interest is the fact that the differences between investigate and legislate ratings were the same for natural kinds and artifacts [$t(17) = 0.38$, n.s.]. In both cases, it was judged more appropriate to resolve a dispute over category membership by discovering additional facts, rather than by making a stipulation. The same pattern held for nominal kinds. The participants preferred investigation to legislation for all category types.

Natural kind categories were judged to be less like matters of convention than were artifact or nominal kind categories. The conclusions about relative differences, however, must be interpreted in light of two important caveats: low absolute levels and item differences in essentialist responding.

The tests presented in Figure 4 indicate that ratings for natural kinds on two measures did not differ significantly from chance (10.5). For the eight natural kind items, an individual was considered to show an essentialist pattern on a given measure if he or she rated no more than one of the items at 10 or lower [p(7 or more of 8) = .02, assuming p(rating > 10) = .5]. Sixteen participants showed this pattern for the investigate measure. However, only 6 and 5 responded according to an essentialist pattern for the legislate and the diversity measures, respectively. Six participants displayed the essentialist pattern for artifacts on the investigate measure, although none did for the other two measures. Note that the criterion of matching a pattern with one or fewer deviations is somewhat more stringent for the artifact than for the natural kind items, because of the larger number of artifacts.

Individual items were analyzed in separate ANOVAs for each type of category (natural kinds, artifacts, and nominal kinds). The data for these analyses were mean essentialism scores (averages of investigate, legislate, and diversity measures) for individual items. No significant item differences were found for the natural kinds [$F(7,119) = 0.647$, $MS_e = 7.0$, n.s.]. There were significant item differences for artifacts [$F(8,136) = 12.0$, $MS_e = 11.6$, $p < .001$] and nominal kinds [$F(6,102) = 3.7$, $MS_e = 15.5$, $p < .01$]. As was discussed above, subtypes within artifacts and nominal kinds had been identified a priori. For artifacts, the scientific instruments received higher essentialism scores than did the other artifacts [$F(8,136) = 11.5$, $MS_e = 11.6$, $p < .001$; Scheffé's test]. These same results can be observed in the pairwise comparisons: The three scientific instruments were each rated significantly higher than all other items. No other pairwise comparisons were significant (Tukey's HSD, $p < .05$). For nominal kinds, mathematical items received higher essentialism scores than did others [$F(6,102) = 2.6$, $MS_e = 15.5$, $p < .05$; Scheffé's test]. The only significant pairwise comparisons (Tukey's HSD, $p < .05$) were between the two mathematical items (TRIANGLE, PRIME) and TAG.

Given the significant variability within category types, it was informative to revisit the category type comparisons. Scientific instruments and natural kinds did not differ significantly on any of the three measures. The largest t value was for the legislate measure [$M_{\text{instruments}} = 9.5$, $M_{\text{natural kinds}} = 11.8$; $t(17) = 1.8$, n.s.]. The only difference between mathematical items and natural kinds came on the same measure: It was judged more acceptable to legislate for mathematical categories [$M_{\text{mathematical}} = 7.9$; $t(17) = 2.6$, $p < .05$].

Thus, at least some artifacts are essentialized to the same degree as are (some) natural kinds. Given the small number of scientific instruments included in the experiment, it is not possible to determine what was responsible for the higher essentialism ratings for these items. Two possibilities suggest themselves. One is that scientific instruments clearly fall under the definitional authority of scientists. For something like a cyclotron, a body of experts has the authority to regulate categorization practices. For more common familiar artifacts, such as socks or chairs, experts' authority may be more circumscribed. Categories may differ in their division of linguistic labor; lay people are more willing to defer to experts on some issues than on others. Future research could explore whether scope of authority is well correlated with judgments taken as evidence of essentialism. Such a correlation could explain the general differences between natural kind and artifact categories. An alternative possibility is that the scientific instruments were essentialized because the physical qualities they measure are essentialized. Thus, people may believe that it is an objective matter of fact whether a quantity is temperature or air pressure, and some thing either measures the quantity or not (all or none). Again, future studies in which a broader range of scientific artifacts is examined could address this possibility.

The issue of the absolute levels of essentialism is a complex one. Based on the ranges of the scales used, the implication of the above experiment seems to be that artifact and nominal kind categories are treated as matters of convention. These items generally received low ratings on measures of essentialism. Ratings for natural kinds were higher. The participants did generally accept empirical investigation as a good strategy for determining membership in natural kinds, although this strategy was preferred for all items. The participants did not strongly reject conventional stipulation of natural kind categories (the *legislate* option). Similarly, people seemed somewhat tolerant of diversity in ascriptions of membership in natural kind categories. These results do not indicate a strong commitment to essentialism. Indeed, the absolute levels of essentialist responding were roughly similar to those observed

in Experiments 1 and 2. Thus, the three experiments may be taken as providing converging evidence that people are essentialists about natural kinds to some intermediate degree.

GENERAL DISCUSSION

Two criteria for essential structure are absolute category membership and objective criteria. In Experiments 1 and 2, the participants did not consistently, or strongly, rate membership in natural kind categories as all or none. Across a range of (potential) natural kind categories, judgments were somewhat intermediate between endorsements of graded and absolute structure. Some categories were seen as having relatively absolute membership. However, it was unclear exactly what distinguished that particular set. In particular, the most commonly discussed natural kind categories, taxonomic kinds of living things, did not receive ratings especially indicative of essential structure. Similar results came from measures assessing beliefs about the objective bases of category identity in Experiments 1, 2, and 3. In general, the participants did not strongly reject the possibility of alternative criteria for determining category identity. Whether an individual is or is not a member of a natural kind category was judged to be a matter of convention or opinion, rather than a matter of objective fact, at least to some degree. Although natural kinds were rated as more objective than were many artifact categories, this was not universally the case. One subtype of artifact categories (scientific instruments) received the same ratings as natural kinds. Thus, the results are quite equivocal. Some natural kinds may be essentialized, but many are not. Some natural kinds are essentialized to a greater degree than some artifacts. The results suggest weak support for essentialism about natural kinds, or support for a weak form of essentialism.

The results of the experiments above are generally consistent with past research on category structure. Kalish (1995) and Hampton (1998) found that membership in natural kind categories was judged to be a matter of degree. In a recent report, Diesendruck and Gelman (1999) have claimed to find evidence to the contrary. However, Diesendruck and Gelman did report a substantial number of natural kind categories receiving less than absolute categorization ratings. Low typicality instances were judged to be category members only to some degree. Since the items in the present study are presented as unusual or difficult-to-decide cases, they also represent atypical instances. Thus, graded membership ratings in the present study are consistent with Diesendruck and Gelman's findings. Diesendruck and Gelman went on to speculate that their participants' graded ratings reflected their appreciation that it was difficult to tell whether instances met the (absolute) criteria for category membership. However, this contention is not empirically supported and seems to be contradicted by the results of the present experiments and by past research (Kalish, 1995). Other evidence in support of the essentialist hypothesis comes from studies of expert categorization. Malt (1990) and Coley and Luhmann (2000) found that people believe experts are in the position to make absolute determinations of category membership for natural kinds. Yet deference to experts is not restricted to categories thought to have real essences. Thus, a willingness to rely on expert authority is consistent with the present findings that natural kinds are seen as somewhat graded and conventional.

The results of this study suggest that comparisons of category structure must be interpreted with caution. There is evidently variability within the classes of categories deemed natural kinds and artifacts. Put differently, the available characterizations of *natural kind* and *artifact* do not identify homogeneous types. It is likely that the distributions of category structure within natural kinds and artifacts overlap to some degree. In fact, this possibility may not be incompatible with existing descriptions of natural kinds. Natural kindhood is often conceptualized as a continuum, with the lines between natural and more artifactual kinds somewhat blurred (Gelman et al., 1994; Keil, 1989; Markman, 1990). Given this graded notion of natural kind and artifact, claims about the structures of such categories must be interpreted in distributional or probabilistic terms. To make such arguments, criteria for sampling are required; to what degree are conclusions about category structure based on a given sample of stimuli characteristic of the population? The results from Experiments 1 and 2 suggest that existing criteria for natural kinds do not yield a set of essentialized categories. The results from Experiment 3 suggest that it is important to do some careful matching when choosing samples of natural kind and artifact categories. For example, differences in the degree to which categories are identified with scientists or experts may affect ratings of category structure. Thus, one implication is that issues of sampling must be addressed in future research on the structure of natural kind and artifact categories.

Natural kinds seem to vary in the degree to which they are thought to have essential structures. Nonetheless, at least of the categories tested, few are strongly essentialized. Similarly, although a few participants may have held strong essentialist beliefs, most displayed only moderate commitment. A question raised by these results is how to make sense of these intermediate responses: What is it to be a partial essentialist? In the remainder of this discussion, I will argue that an *essentialism of degree* is an interpretable and sensible view of category structure. Indeed, such a view, which may be termed *pragmatic*, is similar to the account of scientific categories recommended by several contemporary philosophers (Boyd, 1991; Dupré, 1986).

The basic tenet of a pragmatic view is that categories are constructed to serve purposes, to highlight particular similarities and differences. Since there may be many different motives for categorization, there will be many sets of criteria for forming good categories. Thus, boundaries and criteria for membership will be viewed as relative to the goals and purposes of category users. A pragmatic ap-

proach is similar to Barsalou's (1983) proposal of goal-derived categories. Although goal-derived categories are often associated with ad hoc construction, this need not be the case (see Barsalou, 1993, on the distinction). Another related characterization is sense generation (Braisby et al., 1996; Franks, 1995), which emphasizes that people use categories and word meanings instrumentally to highlight particular sets of features and relationships. A pragmatic approach is also consistent with the claim that categories are embedded in larger knowledge structures, such as theories (Murphy & Medin, 1985). The goal of a theory is to make explanatory and predictive generalizations. Categories, as parts of theories, are also constructions intended to serve explanatory functions.

Not all functions or bases for category construction need be equal. People may recognize better or worse, more or less legitimate, motives for categories. In contemporary Western societies, the purposes of scientists are often privileged. Although for most intents and purposes a TOMATO is best categorized as a VEGETABLE, the scientific considerations that indicate that it is a FRUIT are given considerable weight, even by lay people. Scientific (or other "authorized") standards for categories may be accepted as the best or most generally appropriate. In cases in which the context or goals are not specified (e.g., experimental tasks or linguistic judgments), scientific principles may be the best choices (see Sloman & Ahn, 1999, for a discussion of task differences and category use). Crucially, deference to scientific classification need not carry the implication that one way of categorizing is the true one. Believing that scientists' criteria for establishing categories are the most significant, that ultimately they are the most useful, does not exclude alternatives. If one way of categorizing is true, alternatives must be false. That one way of categorizing is the most useful does not imply that alternatives are useless. Thus, people may be unwilling to endorse absolute or exclusive standards for category membership.

A pragmatic view of categories would seem to be consistent with the results of the experiments reported above. As the possessors of the most useful system of categories, experts have a privileged role in determining category identity. It is generally appropriate to defer to their criteria. Yet such deference does not reflect a belief that there are absolute or objective criteria for category membership. Multiple, partially overlapping criteria for category identity will tend to produce graded membership, as individuals satisfy more or fewer of the criteria (Lakoff, 1986). Criteria for category membership will also be treated as objective to some degree. There are better and worse ways to form categories. Yet, given that interests and motives may vary, alternative systems of categories are not necessarily illegitimate.

In general terms, pragmatic approaches replace questions of truth with questions of justification (Rorty, 1999). The proposal that people are pragmatists with respect to categories suggests that they are less interested in whether it is true that some individual is a member of a particular category than in whether one might be justified in so treating it. Questions such as "Is a TOMATO a FRUIT?" or "Is a SEAHORSE a FISH?" tap intuitions about the motives or utilities of categorization judgments. In determining whether a given way of categorizing is well motivated, intuitive theories and beliefs about experts' criteria are no doubt influential. Sources of justification may be evaluated differently for different categories; scientific opinion may carry more weight for natural kinds than for artifacts, for example. In most cases, however, people seem not to see criteria for categories as completely exclusive. A pragmatic view of categories and the results of the three experiments reported above suggest that people do not see categories as unequivocal. They may be willing to grant experts the final word, but not the only voice.

REFERENCES

Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, **69**, 135-178.

Atran, S. (1987). Folkbiological universals as common sense. In S. Modgil & C. Modgil (Eds.), *Noam Chomsky: Consensus and controversy* (pp. 247-268). London: Falmer.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, **11**, 211-227.

Barsalou, L. W. (1993). Challenging assumptions about concepts. *Cognitive Development*, **8**, 169-180.

Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research*, **18**, 433-447.

Boyd, R. (1991). Realism, anti-foundationalism, and the enthusiasm for natural kinds. *Philosophical Studies*, **61**, 127-148.

Braisby, N., Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. *Cognition*, **59**, 247-274.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, **4**, 73-121.

Coley, J. D., & Luhmann, C. C. (2000). *A watch is sorta clothing but a seahorse is definitely a fish: Typicality, graded category membership, and domain specificity*. Unpublished manuscript, Northeastern University.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin & Review*, **6**, 338-346.

Dupré, J. (1986). Sex, gender, and essence. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy: Vol. 11. Studies in essentialism* (pp. 441-457). Minneapolis: University of Minnesota Press.

Franks, B. (1995). Sense generation: A "quasi-classical" approach to concepts and concept combination. *Cognitive Science*, **19**, 441-505.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, **20**, 65-95.

Gelman, S. A., Coley, J. D., & Gottfried, G. M. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. W. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 341-365). Cambridge: Cambridge University Press.

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

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, **23**, 335-353.

Kalish, C. W. (1998). Natural and artificial kinds: Are children realists or relativists about categories? *Developmental Psychology*, **34**, 376-391.

Kalish, C. W. (in press). Gold, jade, and emeruby: The value of naturalness for theories of concepts and categories. *Journal of Theoretical & Philosophical Psychology*.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kilgarriff, A. (1998). *BNC database and word frequency lists* [Electronic data file]. Available: http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/bnc-readme.html

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 253-355). Dordrecht: Reidel.

LAKOFF, G. (1986). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
LOCKE, J. (1961). *An essay concerning human understanding* (5th ed., Vol. II). New York: E. P. Dutton. (Original work published 1707)
MALT, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory & Language*, **29**, 289-315.
MARKMAN, E. M. (1990). *Categorization and naming in children*. Cambridge, MA: MIT Press.
MEDIN, D. L., & ORTONY, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). New York: Cambridge University Press.
MURPHY, G. L., & MEDIN, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.
RORTY, R. (1999). *Philosophy and social hope*. London: Penguin.
SLOMAN, S. A., & AHN, W.-K. (1999). Feature centrality: Naming versus imagining. *Memory & Cognition*, **27**, 526-537.
SMITH, E. E. (1995). Categories and concepts. In E. Smith & D. Osherson (Eds.), *An invitation to cognitive science: Vol. 3. Thinking* (2nd ed., pp. 3-33). Cambridge, MA: MIT Press.
STREVENS, M. (2000). The essentialist aspect of naive theories. *Cognition*, **74**, 149-175.
TURIEL, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.

NOTES

1. The only label for people included was *human*. This item was judged to have a predominately biological sense. Although this label would seem to denote a taxonomic kind, previous studies of plant and animal kinds have been limited to categories of nonhumans. Thus, to keep the list of taxonomic categories comparable with past research, *human* was not considered a taxonomic item in this experiment. As the results suggest, the category was treated differently than taxonomic terms for nonhumans.

2. Inspection of the results for one of the items chosen to represent a well-defined/absolute category revealed that the item did not perform as expected. The item involved two logicians who disagreed as to whether an argument was logically valid. In contrast to the experimenter's intuitions, the participants rated logical validity as a matter of degree and convention (as akin to such categories as *red* or *friendship*; see Appendix A). Because of this unexpected result, the category *logically valid* was dropped from all analyses comparing well-defined categories with other types.

APPENDIX A
Items Used in Experiments 1 and 2

				Experiment 1			Experiment 2	
Category Type	Expert	Label	Category	Mean, Graded	Mean, Absolute	Mean, Convention	Mean, Objectivity	Mean, Graded
Fuzzy	sociologists	relationship	friendship	4.6	5.1	2.8	11.5	12.2
Fuzzy	present givers	object	gift	2.2	3.0	2.7	2.8	2.4
Fuzzy	artists	color	red	4.8	5.0	4.1	1.6	2.7
	logicians	argument	logically valid	2.9	5.6	5.3	15.4	15.5
Well-defined	mathematicians	number	prime	18.8	19.4	15.2	16.2	17.8
Well-defined	mathematicians	number	sum	15.6	17.5	14.6	15.2	17.6
Well-defined		equations	equal 8		–	–	15.9	16.9
Well-defined		equations	equal each other		–	–	15.9	16.9
Well-defined		hairs	same person		–	–	14.5	17.1
Well-defined		blood	same person		–	–	16.8	16.4
Well-defined		fingerprints	same person		–	–	16.9	18.4
Well-defined		people	same biological father		–	–	15.0	17.1
NB	chemists	substance	water	15.7	15.2	16.4	7.8	8.8
NB	chemists	gas	air	16.5	15.0	14.5	14.6	14.4
NB	chemists	reaction	fire	14.6	15.1	13.8	12.4	12.0
NB	geographers	body of water	sea	8.6	8.7	8.0	7.7	8.9
NB	geologists	substance	oil	12.8	11.7	16.2	7.7	8.5
NB	geologists	substance	stone	7.3	7.6	8.5	8.6	9.7
NB	geographers	body of water	river	10.8	11.1	8.4	8.8	8.1
NB	astronomers	object	star	10.2	11.3	13.7	8.9	9.8
NB	meteorologists	phenomenon	wind	12.2	13.1	9.7	7.6	8.6
NB	geologists	object	rock	7.2	7.6	11.4	8.0	8.4
NB	meteorologists	form of precipitation	rain	8.4	8.7	9.7	11.8	12.9
NB	geologists	formation	mountain	7.6	8.5	9.2	12.7	12.8
NB	chemists	substance	acid	10.9	11.4	13.0	7.9	7.2
NB	oceanographers	phenomenon	wave	8.1	10.2	7.7	15.3	15.6
NB	chemists	substance	metal	10.2	10.0	13.5	13.8	13.7
BNT	biologists	structure	cell	14.5	15.2	16.5	6.7	7.8
BNT	biologists	ecosystem	forest	6.1	7.6	6.2	14.3	15.7
BNT	biologists	plant product	wood	12.1	12.4	12.3	15.0	16.6
BNT	biologists	component of reproductive system	egg	15.6	16.0	15.9	10.4	11.4
BNT	biologists	animal	human	17.6	17.6	16.7	10.4	10.5
Taxonomic	biologists	thing	animal	12.2	12.9	13.5	3.9	5.0
Taxonomic	biologists	plant	tree	9.3	9.7	8.2	12.2	13.0
Taxonomic	biologists	thing	plant	9.0	8.4	12.5	11.5	12.5
Taxonomic	biologists	animal	horse	12.8	12.5	11.3	7.5	8.9

APPENDIX A (Continued)

				Experiment 1			Experiment 2	
Category Type	Expert	Label	Category	Mean, Graded*	Mean, Absolute	Mean, Convention	Mean, Objectivity	Mean, Graded
Taxonomic	biologists	animal	dog	11.9	12.2	10.8	12.5	11.7
Taxonomic	biologists	animal	fish	7.9	8.9	8.5	10.7	12.9
Taxonomic	biologists	animal	bird	9.8	9.9	8.8	9.5	9.2
Taxonomic	biologists	plant structure	flower	6.8	8.2	12.0	10.2	11.0
Taxonomic	biologists	animal	cat	10.5	12.1	11.1	11.4	13.2
Taxonomic	biologists	plant	rose	13.5	13.5	12.4	10.5	10.4
Taxonomic	biologists	animal	sheep	10.8	10.6	11.9	9.0	10.3
Taxonomic	biologists	fruit	apple	12.6	12.1	12.5	10.2	13.2
Taxonomic	biologists	animal	mouse	11.3	11.3	10.9	9.7	10.2
Taxonomic	biologists	vegetable	potato	11.6	12.4	12.3	10.1	11.7
Taxonomic	biologists	animal	chicken	12.5	12.9	13.5	11.4	12.2
Taxonomic	biologists	animal	cow	12.5	13.8	13.3	10.9	12.2
Taxonomic	biologists	animal	pig	11.3	11.8	12.9	10.7	12.6
Taxonomic	biologists	animal	rat	11.5	11.1	10.9	10.3	12.1
Taxonomic	biologists	animal	rabbit	11.5	12.1	11.9	11.2	12.4
Taxonomic	biologists	animal	insect	8.9	9.3	11.1	13.9	15.4

Note—All ratings were recoded so that higher scores are more consistent with essentialism. NB, nonbiological natural kind; BNT, biological nontaxonomic natural kind.

APPENDIX B
Items Used in Experiment 3

Category Type	Category	Mean, Investigate	Mean, Legislate	Mean, Diversity
Artifact	bookcase	11.3	5.4	4.6
Artifact	car	9.8	4.9	5.8
Artifact	chair	8.9	6.3	4.3
Artifact	cyclotron	15.2	9.1	8.9
Artifact	MRI	18.2	9.9	11.1
Artifact	ship	9.4	4.4	5.6
Artifact	shirt	9.4	5.4	4.3
Artifact	socks	8.8	5.4	5.0
Artifact	thermometer	15.6	9.7	10.8
Natural kind	chicken	18.3	11.7	9.2
Natural kind	lemon	15.6	13.1	10.3
Natural kind	marigold	17.4	12.2	10.0
Natural kind	oak	18.0	12.5	10.3
Natural kind	robin	18.0	10.4	10.4
Natural kind	rose	16.9	12.2	9.9
Natural kind	sardine	17.1	9.9	8.4
Natural kind	trout	16.6	12.6	9.4
Nominal	aunt	11.3	9.9	9.4
Nominal	bachelor	10.3	5.9	6.4
Nominal	marriage	11.0	8.0	7.7
Nominal	prime	15.6	6.9	11.6
Nominal	tag	10.7	6.6	4.3
Nominal	triangle	14.4	8.9	10.9
Nominal	triple	11.8	5.1	6.4

(Manuscript received July 27, 2000;
revision accepted for publication December 4, 2001.)