Memory & Cognition
2009, 37 (8), 1150-1163
doi:10.3758/MC.37.8.1150

Feature integration in natural language concepts

James A. Hampton
City University, London, England

Gert Storms
Catholic University of Leuven, Leuven, Belgium

and

Claire L. Simmons and Daniel Heussen
City University, London, England

Two experiments measured the joint influence of three key sets of semantic features on the frequency with which artifacts (Experiment 1) or plants and creatures (Experiment 2) were categorized in familiar categories. For artifacts, current function outweighed both originally intended function and current appearance. For biological kinds, appearance and behavior, an inner biological function, and appearance and behavior of offspring all had similarly strong effects on categorization. The data were analyzed to determine whether an independent cue model or an interactive model best accounted for how the effects of the three feature sets combined. Feature integration was found to be additive for artifacts but interactive for biological kinds. In keeping with this, membership in contrasting artifact categories tended to be superadditive, indicating overlapping categories, whereas for biological kinds, it was subadditive, indicating conceptual gaps between categories. It is argued that the results underline a key domain difference between artifact and biological concepts.

The primary focus of this article is the probabilistic nature of people's categorization of the world around them and the way in which multiple sources of information are integrated to arrive at categorization decisions. A classic study by McCloskey and Glucksberg (1978) demonstrated that across a range of semantic categories, there was both disagreement and inconsistency in classification. Asked to decide whether a pencil is a tool, some people say "yes" and some say "no." When asked the same question again some weeks later, as many as 30% may change their answer. This uncertainty in categorization may even affect experts, as in the case in which animal behaviorists were found to have no clear concept of *behavior* (Levitis, Lidicker, & Freund, 2009).

In part, uncertainty in categorization can be attributed to the multiple dimensions that need to be integrated in arriving at a decision. Most natural categories, such as *chair* or *apple*, are represented with multiple features. A chair has a characteristic appearance, is made of certain materials, can be used in particular ways, and is created by certain processes. An apple has a visual appearance, taste, and texture and has internal biological processes and causal relations to the apple tree on which it grew, the farmer who chose to plant it, and the potential apple trees that may grow from its seeds. The issue to be addressed here is how these different features are combined in order to arrive at a categorization decision. Suppose that a fruit had the appearance of an apple but was picked from a pear tree. How would people resolve this contradictory evidence in deciding whether it is an apple or whether it is a pear? Or consider the lighthouse/bell tower in the small port of Collioure in southwest France (see Figure 1). Beginning in medieval times as a lighthouse to mark the entrance of the port, in the 18th century, with the demise of the port, it was turned into a bell tower for a chapel. How do people decide in this case what kind of thing it is? Is it still a lighthouse? Is it now a bell tower? When a Tuscan-style cupola was added in the 19th century, giving it a more traditional bell tower appearance, did this change its status further (see http://fr.wikipedia.org/wiki/Collioure)?

The aim of our experiments was to investigate how people integrate conflicting information in order to classify objects of this kind. In particular, we were interested in whether each feature has a constant additive effect, independent of the others, or whether the effect of one feature varies as a function of the presence or absence of others.

The integration of information in the mind is a broad issue (Anderson, 1981), which has been investigated in a wide range of situations, ranging from perceptual judgments (Ernst & Banks, 2002) to real-life decision making (Dhami & Harries, 2001). In relation to categorization, a number of different models of feature integration have been proposed (e.g., Reed, 1972). Rosch (1975) suggested that categorization might be related to the number of category features that an item possesses. The greater the

J. A. Hampton, hampton@city.ac.uk

Figure 1. The lighthouse/bell tower in Collioure, southwest France.

similarity of an object or subclass to the category prototype, based on number of matching features, the more likely will be a positive categorization. Counting features is, however, too imprecise a notion to be quantified in the case of natural concepts. Rosch and Mervis (1975) subsequently operationalized the notion of similarity to prototype by developing a procedure for calculating a *family resemblance score*, and this score was shown to predict typicality within a category. In a similar vein, Hampton (1979) suggested that a weighted sum of category features should be used to predict degree of membership, where the degree to which an object matches each category feature, multiplied by the feature's weight for the category concept, is summed across the features of the concept. Feature weights are specific to each concept but are constant across potential objects. Each of these proposals is a version of the independent cue model for categorization (Reed, 1972), according to which each feature or cue lends a constant amount of weight to the decision, independently of the other features.

An alternative to the independent cue model is the set of interactive cue models based on Medin and Schaffer's context model (Medin & Schaffer, 1978; Nosofsky, 1988). Medin and Schaffer proposed that degrees of match for each feature should be multiplied together to determine similarity. In the case of the context model, similarity to a set of stored exemplars was computed, but the same notion can readily be applied to calculating similarity to a prototype representation (Smith & Minda, 1998). Effectively, in a multiplicative similarity model, the weight of each category feature determines the proportional reduction in similarity that results from the removal or replacement of that feature.

Both prototype and exemplar models employ a relatively unsophisticated form of concept representation. In the case of natural concepts (as opposed to simple perceptual categories constructed in the laboratory), an alternative class of models using richer representations has been proposed in which categorization is based on a theoretical understanding or causal model of the world (Ahn & Dennis, 2001; Bloom, 1996, 1998; Murphy, 2002; Murphy & Medin, 1985; Rehder, 2003; Rips, 1989; Sloman, Love, & Ahn, 1998). For example, an object may be categorized as a chair because it was the designer's intention when it was made that it would fulfill a particular role defined in relation to certain cultural situations where sitting occurs (Bloom, 1998; Gelman & Bloom, 2000; Jaswal, 2006; Matan & Carey, 2001; Rips, 1989). Or a creature may be considered to be a tiger because it has some essential property within its cells that causes it to look and behave like one (Gelman, 2003; Medin & Ortony, 1989; Rips, 1989). These accounts of artifact and biological kinds suggest that there is a deeper causal principle that leads to the observed characteristics of an object or organism. Rehder's causal model, for example, proposes that the probability of categorization relies on estimating the probability that the causal model of a given concept would generate the observed characteristics of the object under consideration, including not only the individual features, but also the appropriate pattern of correlation between them. He showed that different estimates of the likelihood of an item's being in a category arise if a particular kind is defined by a causal model in which a common cause leads to three effects, as compared with a model in which three causes lead to a common effect.

Many of these demonstrations of causal reasoning's affecting categorization have relied on carefully constructed novel scenarios with strongly emphasized causal links that participants have to learn during the course of the study. As such, they clearly demonstrate the potential sensitivity of the participants to causal structures. However, they do not, as such, provide evidence that when the everyday world is categorized, such causal information is of key importance. Keil (2003) has pointed out that people frequently overestimate their conceptual competence and that many well-educated people have only the most rudimentary understanding of the causal workings of the world around them. Although the evidence for deeper causal factors' affecting categorization is strong, other studies have also shown that similarity of physical appearance is by no means ignored when either artifacts (Malt & Johnson, 1992) or biological kinds (Hampton, 1995; Hampton, Estes, & Simmons, 2007; Hampton & Simmons, 2003) are categorized. Malt and Sloman (2007) have provided an excellent summary of the different results obtained from research on which aspects of artifacts determine categorization. It remains to be determined whether the best way to account for the pattern of probabilistic categorization is via a model based on the degree of match between the features of the object and the represented concept or via a model that introduces uncertainty through the unreliability with which the defining cause generates the observed features.

There is clearly a diversity of approaches to representing conceptual knowledge. In this article, we choose to focus on a relatively straightforward question. We take it that the most parsimonious account of categorization might be one in which each of a number of different features makes an independent contribution to the categorization decision. Other accounts, either involving a multiplicative formulation of similarity or involving the presence of correlated pairs of features reflecting causal schemas, can be contrasted with this first account by considering the question of whether features combine additively or not. It is to this question that we will now turn.

Feature Integration

In order to determine whether features are being combined additively, a simple test is to look at the influence of a given feature on categorization when other features are either present or absent. If the influence is the same, the independent cue model fits the data best. If the influence is stronger in the presence of other features than in their absence, a nonadditive model, such as a multiplicative model, is needed. In a preliminary examination of this question, Hampton (1995) analyzed the data from four experiments, in each of which six groups of participants categorized six different versions of concept instances. The six versions of an individual concept were constructed using two features. One feature, based on appearance, was either present or absent, whereas the second, based on function (for artifacts) or biological essence (for animals), was fully present, partially present, or absent. When a feature was absent, it was replaced by a contrasting feature of a closely related category. Frequencies of *yes* responses were analyzed to determine whether the effect on categorization probability of changing each feature was independent of or dependent on the state of the other feature. For example, did changing the appearance of a creature have the same size of effect on categorization regardless of whether the essence feature was present, partially present, or absent?

The results indicated that, contrary to the independent cue model, the effects of the two features were not additive. The effect of changing a feature was greater when the other feature was fully present than when it was partially present or fully absent. However, the test was not ideal, involving as it did pooling data from four experiments, with some items repeated between experiments. Nor was it possible to make any differentiation between different conceptual domains. The goal of the present research was, therefore, to examine this phenomenon in a more systematic and controlled fashion. By using three binary features in a $2 \times 2 \times 2$ rather than a 2×3 design, a more comprehensive test of feature independence was possible. The effect of removing any particular feature could be tested at three levels: with both the other features present, with just one other feature present, or with neither present.

We were also interested in the question of whether there may be domain differences in the dependence or independence of features. The issue of how categorization may differ across domains is clearly very important. Wattenmaker (1995) compared social categories (personality traits or occupations) with various object categories, including both animals and artifacts. He found that when people had to categorize a set of items on the basis of features, they were much more likely to rely on counting the number of matching features for social categories than for object categories. He argued that social categories are generally more flexible and more amenable to the resolution of inconsistency. Object categories, on the other hand, may involve interactions among features, rendering some feature combinations more plausible than others. (E.g., to be a "good hammer" it is better for an object to be small and light or large and heavy than to be small and heavy or large and light.) By contrast, in our experiments, we sought to contrast artifact and biological kinds. There is already considerable evidence that these two domains differ in many ways in terms of conceptual structure (Medin, Lynch, & Solomon, 2000). As an early example, Keil (1986) found that transforming the appearance of an artifact would change its type, whereas transforming a biological kind would not. On the other hand, making a discovery about a biological kind could affect its categorization, whereas similar discoveries did not affect artifact categorization. Further work by Kalish (1995) and Estes (2004) showed that many people tend to consider membership in biological kind categories to be an objective fact, so that a disagreement could be resolved by reference to an expert, whereas membership in artifact categories may be more a matter of subjective opinion. There is also interesting evidence of domain-specific semantic aphasia that has been attributed to different patterns of correlation observed among the features of biological and artifact kinds (Cree & McRae, 2003; Tyler, Moss, Durrant-Peatfield, & Levy, 2000). Ruts, Storms, and Hampton (2004) showed similarly that superordinate biological kinds have much tighter similarity clusters than do artifact kinds, allowing the similarity space to be easily divided into linearly separable conceptual categories. We therefore predicted domain differences in how features combine. It has been proposed that biological kinds may be represented with a common cause structure (Gelman, 2003; Rehder, 2003): An underlying genetic cause leads to inner functions and outward appearance and to the appearance of offspring. With a common cause leading to three features, the absence of any one of those features may act strongly to reduce confidence in the cause's still being present. Once this confidence has been reduced, the absence of a second feature will be less critical: The biggest drop in confidence occurs with the first sign of trouble. In contrast, artifact kinds may have a much looser causal structure. The artifact's use and appearance are related to its designer's intentions via a causal path, but we predicted that this influence would be much weaker. The same function can be served by different forms, and the same form can serve different functions. Because membership in artifact kinds is considered to be more a matter of subjective opinion (Estes, 2004; Kalish, 1995), we considered that each feature may have an independent effect on categorization, with people summing the evidence for an object's being of one kind or another in a simple linear additive fashion.

Our first goal was to establish three different aspects of the concepts that would each affect categorization, so that their interaction could be tested. Ideally, the same three aspects would have been selected for each domain. Unfortunately, the domains already differ too much for this to be possible. We therefore aimed not so much to select three individual features as to divide the available semantic information into three parts, on the basis of a commonly understood analysis of the conceptual contents.

For artifacts, there is evidence from both adult and developmental literature that the *function* of an artifact is a highly important feature in determining its class (Bloom, 1998; Malt & Sloman, 2007; Rips, 1989). A central aspect of being a prototypical chair is that it is used for sitting on. In addition to current function, it has been argued that an even more crucial feature is the *function or use for which it was intended* (Bloom, 1996, 1998; Gelman & Bloom, 2000; Jaswal, 2006; Kelemen & Carey, 2007; Matan & Carey, 2001). Thus, the fact that a craftsman designed or constructed an object with the intention that it serve the functions of a chair may override the fact that, for some reason, the object cannot be sat upon or that it is currently in use as a bedside table. A broken chair may still be a chair for this reason. Evidence of the relative importance of current versus historically intended function is mixed. For example, Chaigneau, Barsalou, and Sloman (2004) gave people different scenarios in which various features of an object, such as a mop, were independently manipulated. They found in their studies that a change in current function always had a greater effect on naming than did a change in original intended function. In addition to function, there is also evidence from Malt and Johnson (1992) that the *general appearance* of an object can affect its categorization. For example, a chair is arguably differentiated from a stool more in terms of its appearance (having a back) than in terms of its function (enabling the action of leaning back). Malt and Johnson gave people descriptions of artifacts that had either unusual functions together with a normal appearance or unusual appearance together with a normal function. In this study, participants actually placed more weight on appearance than on function. They confirmed the category membership of objects with normal appearance but unusual functions 58% of the time but those with unusual appearance and normal functions only 25% of the time. This pattern obtained even when the story explicitly stated in the former case that the normal function was not served by the object.

For artifacts, then, the combination of current function, intended function, and appearance was considered to cover all the relevant information.

For biological kinds, there are similarly three major aspects that may, together, be considered crucial to determining type. *Prima facie*, the first aspect is the appearance of a plant or animal. Species may often be differentiated on the basis of their outward physical appearance and behavior. In addition, it has been shown that people entertain essentialist beliefs about biological kinds (Gelman, 2003; Medin & Ortony, 1989). Two kinds of features have been proposed that could figure in such beliefs (Hampton et al., 2007; Strevens, 2000). On the one hand, people may believe that the "innards" of the organism are crucially important (Gelman & Wellman, 1991). If, for example, some biochemical function is present that is characteristically found only within one species, this could be taken as strong evidence that the organism has the essence of that species. On the other hand, people may believe that the genotype constitutes the essence of the organism, so that if the organism has offspring resembling a particular kind, that would constitute strong evidence of the true nature of the parent and, hence, of how the organism should be categorized (Rips, 1989). Hampton et al. obtained evidence using Rips's transformation task that each of these views may be found in a student population.

In our experiments, we therefore aimed to construct materials in which three roughly equally weighted aspects of an object could be independently manipulated. For biological kinds in Experiment 1, they were (1) appearance and behavior of animals or appearance and taste/smell of plants (*appearance*, for short), (2) internal biology (*innards*, for short), and (3) appearance and behavior of the offspring (*offspring*, for short). For artifacts in Experiment 2, these features were (1) appearance, (2) current function, and (3) originally intended function (or *original function*, for short). By forming eight different descriptions of objects, plants, or animals, corresponding to the presence or absence of each of the three features, we were able to measure the relative strength or importance of each type of feature as it affected categorization and then to determine whether they combined in an additive or nonadditive way. In selecting the materials, we were careful to avoid feature combinations that would render an item too implausible. We achieved this by pairing each concept with a similar contrast set—a church paired with an art gallery, a crab with a lobster.

The experiments presented below were scaled-up versions of two experiments presented in Hampton and Simmons (2003). To provide some background on the materials and method to be used, these earlier experiments will be briefly reviewed. Each experiment in Hampton and Simmons used eight pairs of biological kinds (four plants and four animals) and eight pairs of artifacts. Each concept pair consisted of two closely contrasting concepts—for example, *shark* and *whale* or *tie* and *scarf*. For each pair of concepts, three sets of features were identified as above, each with two values—one for the first concept and the other for its contrast. The features of each pair of concepts were combined in all possible combinations to construct eight possible exemplars, which were given to 128 participants to categorize. The participants were given a cover story about a nuclear accident on a large remote island (for biological kinds) or a secluded community in a remote area of Eastern Europe (for artifacts) and were asked to classify each item. The two experiments differed in whether the participants gave a yes/no judgment to each item with respect to one of the categories or whether they chose the category in which the item was best placed.

To summarize the results, biological kinds showed strong and significant influences of all three types of feature on categorization probability. The appearance and behavior of a creature or the appearance and smell or taste of a plant were considered important information for categorization over and above the biological innards and the offspring information. The materials were therefore well suited to the test of cue independence that was planned for Experiment 1 below, since one can test for the moderation of one feature by another only if each feature has a reasonably strong individual effect on the probability of categorization (in the ideal case, the three features would have equally sized effects on categorization). For artifacts, one feature, current function, dominated the rest, in keeping with Chaigneau et al. (2004). In contrast to Malt and Johnson's (1992) study, the appearance of an object had a very small effect on categorization, and in neither experiment did it reach statistical significance. Original function had a minor influence and was statistically significant only in the second experiment. In order to provide materials for a test of cue independence, the artifact concepts for Experiment 2 were therefore adapted in an attempt to balance up the three types of features. Current function was weakened by suggesting that the objects were now only rarely used but that, when they were employed, it was solely with a particular function, and original function was boosted by stating that the object was both designed for that function and used to serve that function in the past. No adjustment was made to the appearance features.

EXPERIMENT 1

Method

Participants. The participants were 375 students at the Catholic University of Leuven, Belgium, who each completed a booklet for course credit.

Design and Materials. The biological kinds used in Hampton and Simmons (2003) were extended and revised. There were 16 pairs of biological kind concepts, half plants and half animals (see Appendix A). The pairs were chosen to be sufficiently similar for a hybrid possessing some features of each to be reasonably plausible (e.g., a crab vs. a lobster). For each pair of concepts, appearance, innards, and offspring features were created. (For simplicity, we refer to the three aspects as *features*, although, in fact, each aspect may be composed of multiple features. Since the aim was to contrast appearance with deep properties, appearance features were taken together as a single set.)

Appearance was a set of features that included behavior for the creatures and either smell or taste, where appropriate, for the plants. *Innards* referred to a biochemical property found in the creature that was specific to only one species. *Offspring* used the same set of appearance features, but they were attributed to the offspring of the organism.

Thirty-two booklets were constructed, each containing a set of instructions to set the scenario and 16 different items to categorize. Illustrative examples may be seen in Appendix B, and the full list of pairs in Appendix A. To illustrate, the first pair of concepts, *crab* and *lobster* (see Appendix B), were used to create a hybrid (appearance = crab, innards = lobster, offspring = crab) by taking the appropriate lines of text from Appendix B to generate the following description and question:

> A creature with legs and claws that looks and acts just like a crab. The scientists found that the structure of its eyes was identical to that typically found only in lobsters. They found that the creature had offspring that looked and acted just like crabs.
>
> Is this a CRAB? YES NO

Half the booklets asked for categorization relative to Concept A (*crab*), and half relative to Concept B (*lobster*) for each pair. Within each booklet, the 16 different items included 2 items for each of the eight possible combinations of the three features. The items were rotated across feature conditions across booklets. Two orders of items in booklets were used. The first order had items randomly ordered within blocks for plants and for creatures, and the second was the reverse of the first. The materials were prepared in English and translated into Flemish Dutch by the second author. The instructions were as follows:

> Many years ago, there was a nuclear accident near to a large remote island populated with a wide variety of animals. The accident resulted in its being contaminated by radiation. At some point in the future scientists are sent to investigate the long-term effects of this accident. They find and examine a number of individual creatures. Can you help them to decide what kind of creature each one is?

(Although, through an oversight, the instructions did not mention plants, in fact, half the biological kinds included were plants. No participants mentioned that they noticed this omission, and since there was no important difference between responses to plants and to animals, it was assumed that they took the instructions to apply equally to all the items in this section of the booklet.)

Results

Feature integration. Booklets were distributed to 384 students, 12 of each of the 32 different booklets in the design, and 375 of these were returned in usable form. In addition to the 9 missing booklets, there were 20 individual missing responses. Overall, missing data accounted for less than 3% of the data. Estimates of categorization probability used in the analyses were based on between 21 and 24 participants per cell. A further exclusion of data was unfortunately required because of a typographic error in the construction of the booklets, which meant that one of the 16 concept pairs (*tiger–wolf*) had to be dropped from the analysis. The results reported below are based on the remaining 30 concepts in 15 concept pairs.

Mean probabilities of categorization are shown in Table 1, together with standard deviations across items. The data for plants and animals were very similar and, so, are reported together. (Animals had a slightly greater effect of appearance than did plants—perhaps because behavior for animals is more salient than smell or taste for plants.) Categorization probabilities were normalized by converting them to z scores for analysis using the inverse of the normal cumulative distribution function with $M = 0$ and $SD = 1$ (see Hampton, 1995, 1998). Thus a probability of .5 was transformed to a z of zero, a probability of .9 became a z of $+1.28$, and a probability of .1 became a z of -1.28. This transformation was justified as follows. In order to account for probabilistic categorization, Hampton (1995) proposed that the weight of evidence for categorization (e.g., the similarity of an item to the concept prototype) is compared with some decision threshold. Both the estimated similarity and the threshold are subject to noise across participants and across occasions. If this noise is assumed to be Gaussian, as similarity to prototype increases, so the probability of categorization will increase in line with the cumulative normal distribution. It follows that a measure of similarity relative to threshold can be calculated from the categorization probability, using the inverse of this function. Effectively, z represents the distance of an item from the category borderline, measured in standard deviations of the variability in categorization across individuals. Positive values of z indicate a greater than 50% chance of being categorized, and negative values a less than 50% chance of being categorized. It is then

Table 1
Means and Standard Deviations Across Items for Probability of Categorization and *z* Scores for Each of the Eight Types of Exemplars in Each Experiment

	Offspring Appearance +								Offspring Appearance −							
	Innards +				Innards −				Innards +				Innards −			
	Appearance +		Appearance −		Appearance +		Appearance −		Appearance +		Appearance −		Appearance +		Appearance −	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Experiment 1: Biological Kinds (*N* = 30)															
Probability	.96	.04	.63	.14	.84	.09	.30	.11	.51	.13	.16	.07	.22	.09	.12	.08
z score	1.77	0.30	0.35	0.38	1.06	0.39	−0.56	0.38	0.03	0.33	−1.04	0.31	−0.83	0.36	−1.24	0.44
	Current Function +								Current Function −							
	Original Function +				Original Function −				Original Function +				Original Function −			
	Appearance +		Appearance −		Appearance +		Appearance −		Appearance +		Appearance −		Appearance +		Appearance −	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Experiment 2: Artifacts (*N* = 31)															
Probability	.95	.07	.84	.13	.79	.12	.59	.18	.41	.14	.30	.17	.11	.09	.07	.08
z score	1.65	0.40	1.11	0.53	0.92	0.51	0.24	0.55	−0.21	0.41	−0.59	0.59	−1.27	0.54	−1.46	0.55

possible to assess how individual features combine in their effect on categorization by looking at how changing one or more features affects the underlying z. The independent cue model predicts that the reduction in z resulting from changing two features should be simply the sum of the effects on z of changing each one individually.

The transformation function gives values of plus and minus infinity for probabilities of 1 or 0. Probabilities of 1 were therefore replaced by 23.5/24, and probabilities of 0 were replaced by 0.5/24 (z of plus and minus 2.04, respectively). Mean and standard deviation (across items) for z are also shown in Table 1. The three features were entered as factors into a three-way repeated measures ANOVA with z as a dependent variable. All three main effects were significant. Mean (and standard deviation) effect sizes (in z) for appearance, innards, and offspring were, respectively, 1.13 (0.28), 0.67 (0.22), and 1.42 (0.26). All main effects had Fs(1,29) greater than 250 (all ps $< .001$). All four interactions were also significant [appearance × innards, $F(1,29) = 4.4$, $p < .05$; appearance × offspring, $F(1,29) = 73.6$, $p < .001$; innards × offspring, $F(1,29) = 22.9$, $p < .001$; three-way interaction, $F(1,29) = 42.1$, $p < .001$]. Breakdown analysis of the three-way interaction showed that the interaction between innards and offspring features was significant when appearance was negative [$F(1,29) = 60.5$, $p < .001$], but not when appearance was positive [$F(1,29) = 2.6$, $p > .10$]. Similarly, appearance interacted with innards when offspring was negative [$F(1,29) = 25.9$, $p < .001$], but not when offspring was positive [$F(1,29) = 2.7$, $p > .10$]. All of the interactions took the form of reduced effectiveness of one feature when the other was missing.

The pattern of significant interactions was consistent with a mode of combination of features in which a feature has more weight in the presence of other features. The independent cue model, as adopted by early prototype models (Hampton, 1979; Reed, 1972), would predict that features contribute equally to similarity, independently of each other. The results suggest a dependence between the features, as would be found in a multiplicative model where the degrees of mismatch for each feature are multiplied together to determine dissimilarity (Medin & Schaffer, 1978; Smith & Minda, 1998). The result similarly supports causal models that effectively give weight to the presence of the predicted correlations among features (Rehder, 2003).

A second analysis directly compared the effect of removing a single feature (measured in z-transformed probability) when both other features were present, when only one was present, or when neither was present. The result is shown in the top panel of Figure 2. For all three features, the effect of removing the feature was greater when other features were present than when they were absent. There was no difference, however, between the case in which just one other feature was present and the case in which both were present. The pattern was confirmed with a significant effect of feature presence overall [$F(1.6, 47) = 49.3$, $p < .001$, with Greenhouse–Geisser correction] and for each individual feature (all Fs > 20.2, ps $< .001$). In each case, both linear and quadratic contrasts were significant. For the analyses of overall feature presence and of the individual features of appearance, innards, and offspring, linear contrasts had Fs(1,29) of 52.3, 46.1, 13.6, and 80.2, respectively (all ps $< .001$). Equivalently, all four quadratic contrasts had Fs(1,29) of 42.1 (ps $< .001$).

A final check was run on whether the interaction effects could, in part, be the result of a high level of positive responses to the [−−−] stimulus (surprisingly, there were 12% *yes* responses to this set of stimuli). Ten of the 30 concepts were selected with the constraint that the [+++] stimulus had at least 90% *yes* responses and the [−−−] stimulus had at least 90% *no* responses. The resulting data looked very similar, with a mean change in z of 1.43 when both features were present, 1.25 when just one was present, and 0.72 when neither was present. The main effect of feature presence was significant [$F(2,18) = 14.9$, $p < .001$], again with a strong linear trend [$F(1,9) = 23.8$, $p < .001$].

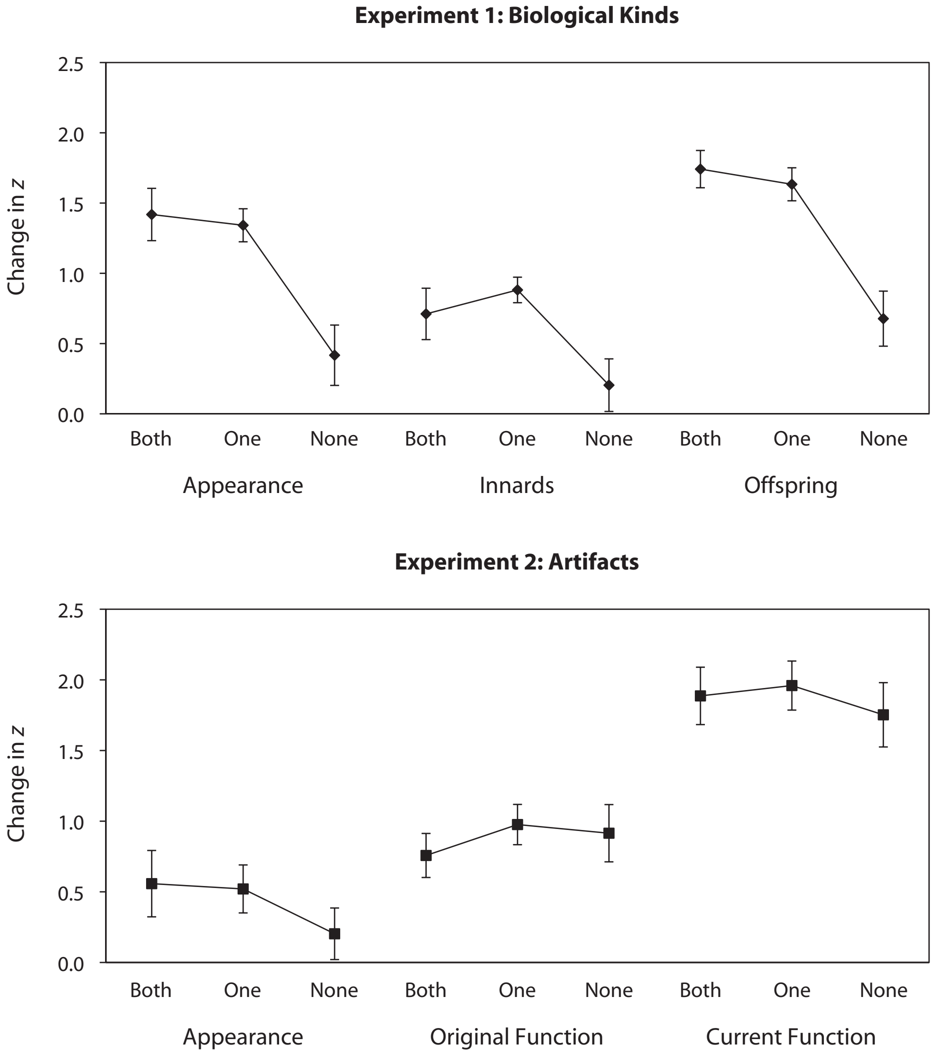

Figure 2. Effect of changing each feature in Experiment 1 (biological kinds) and Experiment 2 (artifacts) on the *z*-transformed probability of categorization when both, just one, or neither of the other two features was present.

Truth gaps and gluts. Having the same stimuli categorized by half the participants for one concept (e.g., *crab*) and by the other half for the other (contrasting) concept (e.g., *lobster*) meant that it was possible to determine the degree to which the two probabilities sum to one. If features are combined multiplicatively, one should expect a stimulus that has some features of each concept to fall into a *conceptual gap* between the two concepts. For example, a creature that had some crab features and some lobster features may tend to be considered *neither* a crab *nor* a lobster. To test for this, the observed probabilities of a creature's being classified in either category were summed for each type of stimulus. Since the [+−+] stimulus for *crab* was the [−+−] stimulus for *lobster* (and so forth), the eight stimuli could be paired up into four possible conditions.

The interesting cases were those in which a stimulus combined one feature of one concept with two features of the other. These creatures or plants were chimerical, having properties of more than one type. These cases were broken down into three kinds, according to which one of the features was pitted against the other two. When either appearance or offspring was at odds with the other features, the sum of the two alternative categorizations was, in each case, significantly below 1 [$M = 0.85$, $SE = 0.03$, $t(29) = 5.1$, and $M = 0.81$, $SE = 0.03$, $t(29) = 6.6$, respectively; $p < .001$ in each case]. There was a truth *gap*. Creatures or plants with inconsistent feature combinations were more likely to be rejected from both classes than included in both. (The pattern for plants and animals did not differ significantly.)

Surprisingly, the final set of chimerical cases—those in which innards were opposed to appearance and offspring—did not show subadditivity. The sum of probabilities was 0.999 ($SE = 0.02$), which was clearly not different from 1. When the hidden biological function was at odds with the observable facts about the creature and its offspring, the degree to which the creature was categorized in one category was exactly matched by the degree to which it was *not* categorized in the other.

Discussion

Experiment 1 provided a powerful test of the way in which features are integrated in determining categorization for biological kind concepts. Thirty biological concepts were tested, with more than 20 participants categorizing each of the eight stimuli for each concept. The results confirmed the significant part played by all three features in categorization of biological kinds. Appearance, innards, and offspring all affected the likelihood of categorization. In addition, the results confirmed that features of biological kinds are combined in an interactive way, consistent, for example, with either a multiplicative rule or a common cause model. The effect of changing any of the features into that of its contrasting concept was greater when the other features were present and was much lower when both others were missing. As a result, creatures or plants with inconsistent sets of features tended to fall between the two categories and were more likely to belong to neither kind than to belong to both.

Interestingly, the drop in effect size with other features was not linear in Figure 2. There was little change between both and one feature present and then a large drop in effect size when both were absent. When both other features are absent, a feature is on its own and in a minority. Although this attracts a certain number of *yes* responses, its influence on responding is quite small. When one other feature is present and the other is absent, however, the feature in question holds the "deciding vote," turning the number of matching features from a minority to a majority, and hence, the feature's influence is much greater. Finally, when both other features are present, the feature in question is the first feature to show that the organism is odd in some way. It is here that the large effect size indicates that feature integration is nonadditive. Even though the two other features still hold the majority vote, the effect of losing the first of three features is large and is much greater than the effect of losing the last of three features.

It is also interesting to note that the results are inconsistent with a simple *counting* strategy of the kind identified by Wattenmaker (1995) for social categorization. If people simply categorized according to the majority of features, the strongest effect of changing a feature would be seen in the case in which that feature holds the deciding vote. The fact that the effect of a feature change was still strong even when the other two features had already determined the majority vote shows that counting was not a commonly used strategy.

EXPERIMENT 2

Experiment 2 followed the same design as the first experiment but used artifact concepts. Prior research by Hampton and Simmons (2003) failed to show convincing evidence that appearance was considered important in categorizing artifacts once original and current functions were defined. In order to test the interaction between features, it was therefore necessary to adapt the materials in order to balance up the features. Appearance was left as it was, but current function was downplayed by stating that the object was now only rarely used but that, when it was used, it had only that function. At the same time, original function was strengthened by stating not only that the object was designed to serve a given function, but also that it did originally serve that function. Of course, if appearance is truly irrelevant to artifact categorization, we would not expect it to have an influence on categorization here. However, given the earlier results of Malt and Johnson (1992), there was reason to suppose that all three features would, in fact, affect categorization, enabling additivity to be tested.

Method

Participants. The participants were 320 students at the Catholic University of Leuven, who participated for course credit. No booklets were returned incomplete.

Design and Materials. The preparation of materials followed exactly the same design as that in Experiment 1, except that 16 pairs of artifact concepts were used in place of 16 pairs of biological kinds, and the three features manipulated were appearance, original function, and current function. Pairs of similar concepts were chosen so that the features could plausibly be swapped between them (see Appendix B for examples, and Appendix A for the full list of pairs). Materials were prepared in English and translated into Flemish Dutch by the second author. Note that original function and current function were deliberately the same in the case in which the two features were both positive (or both negative), since the normal scenario for an object is that its current function matches its intended function. In order to reduce the dominant strength of current function as a feature, the current function of the objects was made to sound occasional. For example, the putative church was "occasionally used for Christian services, and has no other function," or the putative tie was described as follows: "now, when used at all, it is only ever worn with shirts and suits by male members of the community as a part of formal dress." Instructions included a scenario intended to lend some degree of plausibility, as follows:

> Anthropologists visited a secluded community in a remote area of Eastern Europe, where they found and studied a number of cultural artifacts. The members of the community were very resourceful and had found ways of sometimes adapting things to new uses. The anthropologists were puzzled about how each item should be classified. Can you help them to decide what kind of thing each one is?

As before, the categorization question was "Is this an A?" for half the participants and "Is this a B?" for the other half, using the category names listed in the Concept A and Concept B columns in the Appendices.

Results

Feature integration. Probability of categorization for each stimulus in each of the 32 concepts was estimated from the frequency of *yes* responses, and the results are presented in Table 1, along with corresponding *z* score data. Examination of individual pairs of concepts revealed an unanticipated effect in the case of the concept pair *theatre*/*cinema*. In Flemish (as in U.S. English), a cinema may also be called a theater, so that even when the object had only cinema features, it was still categorized as a theater by 70% of the participants. (In the U.K., the term *theatre* means primarily a place for live performances of plays.) The item *theatre* was therefore excluded from the analysis. The contrasting category (*cinema*) could, however, still be used, since theaters used for live plays are never called cinemas in Belgium.

Frequencies for the remaining 31 concepts were converted to z scores. Since there were 20 responses per probability estimate, probability values of 0 were taken as 0.5/20, and values of 1 as 19.5/20. The z scores were submitted to a three-way repeated measures ANOVA across items, with features as factors. All three features had strongly significant effects on categorization frequency. Main effects (and standard errors) in z score differences and their F ratios were the following: appearance, $z = 0.45$ (0.07), $F(1,30) = 44.2$; original function, $z = 0.91$ (0.06), $F(1,30) = 238.8$; and current function, $z = 1.89$ (0.07), $F(1,30) = 698.1$ (all ps $< .001$). In contrast to the biological kinds in Experiment 1, there was less evidence that the features interacted. Original function did not interact significantly with current function [$F(1,30) = 3.56$, $p = .07$] or with appearance ($F < 1$), but current function and appearance did interact [$F(1,30) = 10.6$, $p < .005$]. As in earlier experiments, the interaction showed that one feature had a greater effect when the other was present than when it was absent. The three-way interaction was marginal but not significant [$F(1,30) = 3.42$, $p = .07$].

As in Experiment 1, the effect of removing one feature (i.e., replacing it with that of the contrasting concept) was measured when both, just one, or neither of the other features were present. The results are displayed in the lower panel of Figure 2. Unlike the biological kinds, there was no clear trend for the changing of a feature to have a greater effect when the other features were present. Degree of change in z was entered into a two-way ANOVA with feature (three levels) and presence/absence of the other features (three levels) as repeated measures factors across the 31 concepts. There was no overall main effect of presence/absence of other features [$F(2,60) = 1.89$, $p = .16$], but there was an interaction of this factor with type of feature [$F(1.8, 53) = 7.8$, $p < .005$, with Greenhouse–Geisser correction for significant lack of sphericity]. The interaction can be seen in Figure 2 (lower panel). Breakdown analysis of the interaction confirmed that appearance was the only factor that had reduced effectiveness when the other factors were absent [linear trend $F(1,31) = 5.8$, $p < .05$]. Neither of the other two features showed any effect of presence/absence of other features when considered alone [linear trend $F(1,30) = 2.1$ for original function and $F < 1$ for current function; both ps $> .15$].

Finally, given the very different pattern of results from Experiments 1 and 2, a direct comparison was made between them (the design, language, and participant populations were the same in both experiments). Since the features did not correspond between domains (with the exception of appearance), the three features were collapsed within each domain, and an ANOVA was run across the 61 items (30 biological kinds and 31 artifacts), with presence of other features as a repeated measures factor with three levels (both present, just one present, neither present) and domain (experiment) as a between-items factor with two levels (artifacts and biological kinds). The effect of interest was the interaction between feature presence and domain, and this interaction proved highly significant [$F(1.7, 102) = 16.6$, $p < .001$, with Greenhouse–Geisser correction for sphericity].

Truth gaps and gluts. In Experiment 1, there was significant subadditivity when the likelihood of an inconsistent stimulus's being in one biological kind category was added to the likelihood of its being in the contrasting category. A similar analysis was conducted for the artifacts in Experiment 2. The observed probabilities of an object's being classified in either category were summed for each pair of concepts to yield four summed probabilities.

Hybrid cases in which a stimulus combined one feature of one concept with two features of the other were again broken down according to which one of the features was pitted against the other two. When appearance was at odds with current and original function, or when current function was at odds with the other two features, categorization was still additive, with summed probabilities of 0.95 ($SE = 0.03$) and 0.99 ($SE = 0.03$), respectively, not significantly less than 1 [$t(30) = 1.86$, $p = .07$, and $t(30) = 0.4$, $p > .5$]. However, when original function contradicted appearance and current function, there was a significant tendency for categorization to be *superadditive*, with a summed probability of 1.09 ($SE = 0.03$) [$t(30) = 3.37$, $p < .005$]. This result was in stark contrast to the biological kinds, where the general trend was for categorization to be *subadditive*.

Whereas the biological kinds in Experiment 1 had tended toward a truth gap, so that items falling between two concepts were likely to be considered not to belong to either, in Experiment 2 items lying between two artifact concepts showed no truth gaps. Likelihood of being in one category was well predicted by the likelihood of *not* being in its contrast, and where the data deviated from this pattern, items with mixed features were more likely to be categorized in *both* categories than to be placed in neither—a so-called *truth glut* (Bonini, Osherson, Viale, & Williamson, 1999).

Discussion

The results of Experiment 2 contrasted strongly with those from the first experiment. With minor exceptions, the effect of changing one of the features of an object—be it the original function, the current function, or the appearance—was equivalent, regardless of the other properties of the object. As a consequence, there were no truth gaps between concepts. Indeed, there was some evidence for objects' falling into more than one class at the same time. As was discussed above, this result is consistent with other evidence that artifact categorization is based on underlying causal schemas much less than is the case for biological kinds.

It is also interesting to note that all three types of feature had a role to play in categorization. Given the strong advocacy of functions as the basis of defining artifact types, it was interesting that an object with the wrong appearance was not as well accepted as one with the correct appearance, even when both original and current functions were the same. Probability of categorization decreased from 95% to 84% when the appearance mismatched the category. So, at least for some items and some participants, appearance was enough to overrule function. Of the 31 items, 22 had reduced categorization probability when appearance was the only mismatching feature.

GENERAL DISCUSSION

The major question driving the research concerned the way in which features are integrated when category membership is judged. Two possibilities were considered: that features contribute *independently* to the similarity of an instance to a category concept (and hence, its probability of categorization) and that features *interact* in their effect. Our results demonstrated very clearly that different systems appear to be in operation for artifact and biological kind categories. For artifact categories, the effect on categorization of altering a feature was broadly the same regardless of whether the other two features were present or absent. In line with this result, there was very little interaction between features in the ANOVA, and when an instance lay between two different categories (having some features of each), the probability of being in both categories (as measured independently) slightly exceeded 1. In contrast, biological kind categories showed an interactive pattern of feature integration. The effect of altering a feature was much greater when others were present than when both were absent, and all four interactions were significant in an ANOVA. As a consequence, the likelihood of an intermediate instance's falling in both of two contrasting categories tended to be less than 1.

The different pattern of integration for artifacts and biological kinds is perhaps the most important result from this research. Discussion of categorization models in the literature has tended to assume that one model for relating feature possession to categorization should fit all cases (Nosofsky, 1988; Smith & Minda, 1998). Although there has been much discussion of important differences between artifact and biological kind domains (e.g., Estes, 2004; Gelman, 2003; Kalish, 1995; Keil, 1986), this is the first clear demonstration that the way in which information is integrated in these two general domains is different. Wattenmaker (1995) showed a differentiation between social and object categories in terms of the likelihood of generating and the ease of learning linearly separable classifications. However, he did not directly address the question of feature independence or the relation between biological and artifact kinds. On the basis of our findings, we would predict that social and personality concepts should show results akin to those of artifacts. Each aspect of the concept should contribute independently to classification, and it should be more likely that someone with a mixed set of features will belong in both contrasting categories than that they will belong in neither. (One might further speculate that social categories with a higher level of essentialist beliefs will be more like biological kinds; see Haslam, Rothschild, & Ernst, 2000.)

With the benefit of hindsight, it is not difficult to provide some plausible accounts of the difference we have observed. Artifact kinds lack a strong underlying network of causally linked properties. In fact, Sloman and Malt (2003) have argued that artifact categories are not true conceptual categories at all but correspond more closely to *naming* categories—items that, for one reason or another, have happened to end up with the same name. The loose and overlapping landscape of artifact categories lends itself readily to the notion that instances that fall between two categories could be considered to be in both, rather than in neither, just as objects can exist with multiple functions (such as the camera–phone or the fax–printer–copier). (For a wider discussion of the metaphysics and psychological representation of artifacts, see Margolis & Laurence, 2007.)

On the other hand, our beliefs about biological kinds tend to include the notion that there is a strong set of causal principles within each organism that lead to the homogeneity of the class as a whole (Rehder, 2003). Boyd (1999) referred to this notion as *causal homeostasis*. At least in folk understanding of biological kinds, the classes represent tight clusters of similar items with large gaps in between (Ruts et al., 2004). Even relatively close categories, such as foxes, wolves, and husky dogs, are assumed to form easily distinguishable categories in terms of their appearance, and it is assumed that underlying the similarity of appearance is some deeper causal story involving innards and germ lines. In this domain, it makes sense that some individual creature that had the appearance of one type of animal but the innards of another should be considered to belong to *neither* category, rather than to both (just as a cross between two breeds of dog belongs to neither breed, or a mule is considered neither a horse nor a donkey). The interactive pattern of feature integration found for biological kinds reflects the integrated nature of the features. The first feature to be altered (be it appearance, innards, or offspring) immediately casts doubt on whether the organism has the full set of interlocking features that characterize a "proper" member of the kind. The effect of an altered feature is therefore greatest when the others are present.

In order to conduct the test of additivity, it was first necessary to find different sets of semantic features that determine the likelihood of an instance's being placed in a class. Although relative weights cannot be directly compared (feature sets not having been randomly sampled, and varying in many other ways), the fact that it was possible to show main effects of all three sets of features on categorization probability was an important result. The finding that original function is not the only factor affecting artifact categorization may, at first glance, appear to contradict the position advocated by theorists such as Bloom (1996, 1998), who have argued that the kind of an artifact is determined exclusively by its creator's intention. If a designer had it in mind to create a chair, it should not matter what the object looks like or whether it can be sat upon; it is still a chair. In contrast, we found that the original intended function of our artifacts played a relatively minor role in determining categorization (twice the effect size of appearance, but only half that of current function). A way to resolve the issue would be to amend Bloom's thesis. In our covering story, we explained how objects had been adapted to new purposes, and it would therefore seem plausible that the kind of an object is not determined by its *first* designer, but by its *most recent* designer. When the people in our story took a church and started to use it for art exhibits (so that it no longer was used as a church), they were thereby "rebaptizing" the object as a new kind. This interpretation would also fit with Keil's study of transformations of artifacts (Keil, 1986). When children were told of how a metal coffeepot was reshaped and hammered so it looked like and could be used as a bird feeder, they were happy to allow that it was no longer a coffeepot. The majority of our participants would appear to agree. Nonetheless, even when an object was categorized by current function, the appearance of the object continued to have an effect.

For biological kinds, a common view is that categorization is driven by some notion of *essence* (Gelman, 2003; Murphy, 2002). Hampton et al. (2007) argued that there are two discrete notions of essence. One relates to the causal processes at work deep within an organism that lead to its appearance and behavior, and the other relates to the notion of a germ line that is passed from parent to offspring. In our scenarios, each of these two types of information was available to indicate whether the essence was that of the category in question. The data were very clear in indicating that not only both of these more essentialist criteria, but also the appearance and behavior of the organism itself were all treated as relevant sources of information for categorization. There was no clear "winner" in terms of "innards" essentialism versus "germ line" essentialism or, for that matter, outward appearance features. Our choice of features could be criticized by arguing that appearance ("looks and acts just like a crab") provides evidence about deeper essences and that "innards" features such as eye structure or blood chemistry are not themselves particularly deep causal features and, therefore, may be less essential. We do not, therefore, make any strong claims about how the different aspects of biological kinds are used to make category judgments, other than the conclusion that multiple sources of information are involved and the integration of information is interactive (nonadditive) in the case of this domain.

Conclusion

The results presented here constitute the first demonstration of a key difference between biological and artifact kinds. For biological kinds, the evidence for kind-ship is integrated in a nonadditive fashion, with the result that a chimerical creature with aspects of more than one species is more likely to be classified in neither than in both. For artifact kinds, features are apparently combined in an additive fashion, and items with hybrid features may be more likely to be in both categories than in neither. Returning to the problem of the lighthouse/bell tower in Collioure, a Google search on August 28, 2008 gave 17,800 hits for "Collioure phare" (lighthouse) and 21,900 hits for "Collioure clocher" (bell tower). Current function has it by a short whisker.

AUTHOR NOTE

We acknowledge the helpful comments of contributors to the 3rd Leuven Workshop on Concepts and Categories, held in Leuven, Belgium in 2006. Inquiries concerning this article should be addressed to J. A. Hampton, Department of Psychology, City University, Northampton Square, London EC1V OHB, England (e-mail: hampton@city.ac.uk).

REFERENCES

AHN, W. K., & DENNIS, M. J. (2001). Dissociation between categorization and similarity judgment: Differential effect of causal status on feature weights. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 87-107). New York: Oxford University Press.

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

BLOOM, P. (1996). Intention, history, and artifact concepts. *Cognition*, **60**, 1-29. doi:10.1016/0010-0277(95)00699-0

BLOOM, P. (1998). Theories of artifact categorization. *Cognition*, **66**, 87-93. doi:10.1016/S0010-0277(98)00003-1

BONINI, N., OSHERSON, D. N., VIALE, R., & WILLIAMSON, T. (1999). On the psychology of vague predicates. *Mind & Language*, **14**, 377-393. doi:10.1111/1468-0017.00117

BOYD, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141-186). Cambridge, MA: MIT Press.

CHAIGNEAU, S. E., BARSALOU, L. W., & SLOMAN, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General*, **133**, 601-625. doi:10.1037/0096-3445.133.4.601

CREE, G. S., & MCRAE, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *chipmunk*, *cherry*, *chisel*, *cheese*, and *cello* (and many other such concrete nouns). *Journal of Experimental Psychology: General*, **132**, 163-201. doi:10.1037/0096-3445.132.2.163

DHAMI, M. K., & HARRIES, C. (2001). Fast and frugal versus regression models of human judgement. *Thinking & Reasoning*, **7**, 5-27. doi:10.1080/13546780042000019

ERNST, M. O., & BANKS, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433. doi:10.1038/415429a

ESTES, Z. (2004). Confidence and gradedness in semantic categorization: Definitely somewhat artifactual, maybe absolutely natural. *Psychonomic Bulletin & Review*, **11**, 1041-1047.

GELMAN, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

GELMAN, S. A., & BLOOM, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, **76**, 91-103. doi:10.1016/S0010-0277(00)00071-8

GELMAN, S. A., & WELLMAN, H. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, **38**, 213-244. doi:10.1016/0010-0277(91)90007-Q

HAMPTON, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **18**, 441-461. doi:10.1016/S0022-5371(79)90246-9

HAMPTON, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory & Language*, **34**, 686-708. doi:10.1006/jmla.1995.1031

HAMPTON, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, **65**, 137-165. doi:10.1016/S0010-0277(97)00042-5

HAMPTON, J. A., ESTES, Z., & SIMMONS, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition*, **35**, 1785-1800.

HAMPTON, J. A., & SIMMONS, C. L. (2003, November). *Feature independence in natural categories*. Poster presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver.

HASLAM, N., ROTHSCHILD, L., & ERNST, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, **39**, 113-127. doi:10.1348/014466600164363

JASWAL, V. K. (2006). Preschoolers favor the creator's label when reasoning about an artifact's function. *Cognition*, **99**, B83-B92. doi:10.1016/j.cognition.2005.07.006

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

KEIL, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition: Foundational issues* (pp. 133-153). Norwood, NJ: Ablex.

KEIL, F. C. (2003). Categorization, causation, and the limits of understanding. In H. E. Moss & J. A. Hampton (Eds.), *Conceptual representation* (pp. 663-692). Hove, U.K.: Psychology Press.

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 212-230). Oxford: Oxford University Press.

Levitis, D. A., Lidicker, W. Z., Jr., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour*, **78**, 103-110. doi:10.1016/j.anbehav.2009.03.018

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory & Language*, **31**, 195-217. doi:10.1016/0749-596X(92)90011-L

Malt, B. C., & Sloman, S. A. (2007). Artifact categorization: The good, the bad, and the ugly. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 85-123). Oxford: Oxford University Press.

Margolis, E., & Laurence, S. (Eds.) (2007). *Creations of the mind: Theories of artifacts and their representation*. Oxford: Oxford University Press.

Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition*, **78**, 1-26. doi:10.1016/S0010-0277(00)00094-9

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, **6**, 462-472.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology*, **51**, 121-147. doi:10.1146/annurev.psych.51.1.121

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238. doi:10.1037/0033-295X.85.3.207

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

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316. doi:10.1037/0033-295X.92.3.289

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 700-708. doi:10.1037/0278-7393.14.4.700

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407. doi:10.1016/0010-0285(72)90014-X

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, **27**, 709-748. doi:10.1016/S0364-0213(03)00068-5

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rosch, E. R. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-232. doi:10.1037/0096-3445.104.3.192

Rosch, E. R., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605. doi:10.1016/0010-0285(75)90024-9

Ruts, W., Storms, G., & Hampton, J. [A.] (2004). Linear separability in superordinate natural language concepts. *Memory & Cognition*, **32**, 83-95.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228. doi:10.1016/S0364-0213(99)80039-1

Sloman, S. A., & Malt, B. C. (2003). Artifacts are not ascribed essences, nor are they treated as belonging to kinds. In H. E. Moss & J. A. Hampton (Eds.), *Conceptual representation* (pp. 563-582). Hove, U.K.: Psychology Press.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436. doi:10.1037/0278-7393.24.6.1411

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition*, **74**, 149-175. doi:10.1016/S0010-0277(99)00071-2

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain & Language*, **75**, 195-231. doi:10.1006/brin.2000.2353

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology*, **28**, 274-328. doi:10.1006/cogp.1995.1007

APPENDIX A
Full Set of Concept Pairs Used in the Experiments

Biological Kinds (Experiment 1)		Artifacts (Experiment 2)	
Concept A	Concept B	Concept A	Concept B
crab	lobster	church	gallery
mosquito	wasp	banknote	stamp
pigeon	crow	TV	monitor
lizard	snake	cinema	theatre
tiger	wolf	ferryboat	warship
shark	dolphin	taxi	ambulance
rabbit	squirrel	tie	scarf
horse	cow	vase	carafe
rose	dandelion	saucepan	helmet
mint	onion	beer glass	jar
oak	pine	sketchbook	diary
grass	moss	roof tile	drain cover
grape	cherry	drum	waste basket
apple	orange	nightshirt	dress
pumpkin	watermelon	rug	blanket
carrot	potato	chimneypot	flowerpot

(*Continued on next page*)

APPENDIX B
Sample of the Materials Used in Experiments 1 (Biological Kinds) and 2 (Artifacts)

Concept A	Concept B	Feature Type	Concept A Feature	Concept B Feature
			Biological Kinds	
Crab	Lobster	Appearance	A creature with legs and claws that looks and acts just like a crab	A creature with a long tail and claws that looks and acts just like a lobster
		Innards	The scientists found that the structure of its eyes was identical to that typically found only in crabs.	The scientists found that the structure of its eyes was identical to that typically found only in lobsters.
		Offspring	They found that the creature had offspring that looked and acted just like crabs.	They found that the creature had offspring that looked and acted just like lobsters.
Mosquito	Wasp	Appearance	A small flying insect with transparent wings that bites people, and looks and acts just like a mosquito.	A striped flying insect that stings people, and looks and acts just like a wasp.
		Innards	The scientists found that the chemistry of its blood was just like that normally only found in mosquitoes.	The scientists found that the chemistry of its blood was just like that normally only found in wasps.
		Offspring	They found that the eggs laid by the creature developed into offspring that looked and acted just like mosquitoes.	They found that the eggs laid by the creature developed into offspring that looked and acted just like wasps.
Oak	Pine	Appearance	A tall tree that loses its leaves in winter and that looks just like an oak	A tall tree that keeps its needles all year round and that looks just like a pine
		Innards	The scientists found that the microstructure of the wood fibers was just like that only typically found in oaks.	The scientists found that the microstructure of the wood fibers was just like that only typically found in pines.
		Offspring	They found that when the tree reproduced, new trees grew that looked just like oaks.	They found that when the tree reproduced, new trees grew that looked just like pines.
Grape	Cherry	Appearance	A small round green fruit which looks and tastes just like a grape	A small, dark red fruit, which looks and tastes just like a cherry
		Innards	The scientists found that its cellular potassium metabolism is just like that which is normally only observed in grapes.	The scientists found that its cellular potassium metabolism is just like that which is normally only observed in cherries.
		Offspring	They found that when the seeds of this fruit are planted, a vine grows, yielding fruit which look and taste just like grapes.	They found that when the seeds of this fruit are planted, a tree grows, yielding fruit which look and taste just like cherries.
			Artifacts	
Church	Art gallery	Appearance	A large building with stained glass windows, and a steeple with a cross on the top, which looks just like a church	A large gothic building with white interior walls on which paintings are hung, and which looks just like an art gallery
		Original function	It was originally built just to be a place of Christian worship, and had that function in the past.	It was originally built just to be an exhibition hall for displaying large works of art, and had that function in the past.
		Current function	It is presently occasionally used for Christian services, and has no other function.	It is presently occasionally used for the public exhibition of painting and sculpture, and has no other function.
Banknotes	Postage stamps	Appearance	Rectangular pieces of paper with a colored design and an embedded metallic strip which look just like bank notes	Small rectangular pieces of paper with serrated edges and sticky backs which look just like postage stamps
		Original function	Originally, these were produced as a kind of money, and they served that function in the past.	Originally, these were produced for sticking on letters as postage, and they served that function in the past.
		Current function	They are now *just* sometimes used for buying or selling things, and have no other use.	They are now *just* sometimes stuck to envelopes to pay for postage, and have no other use.

APPENDIX B (Continued)

Concept A	Concept B	Feature Type	Concept A Feature	Concept B Feature
Taxi	Ambulance	Appearance	A motor vehicle which is black with a yellow light on the top, has a diesel engine and looks just like a London taxi.	A motor vehicle which is white and green with a flashing blue light on the top, and which looks just like an ambulance.
		Original function	It was originally intended and used to provide transport for small groups of people to their desired destination.	It was originally intended and used to carry sick or injured people to hospital for urgent medical attention.
		Current function	Now, when it is used, its only use is to take people wherever they want to go in exchange for money.	Now, when it is used, people use it *only* in the case of medical emergencies when the driver takes people to the hospital
Tie	Scarf	Appearance	An item sewn from a long piece of patterned silk fabric which looks just like a man's tie	An item which is made of a long thin piece of knitted wool, and looks just like a scarf
		Original function	Originally, it was intended to be tied around the collar of a shirt as a form of decoration, and in the past it had this function.	Originally, it was intended to be wrapped around the neck for protection against the cold when outside, and in the past it had this function.
		Current function	Now, when used at all, it is *only ever* worn with shirts and suits by male members of the community as a part of formal dress.	Now, when used at all, it is *only ever* worn round the neck by members of the community for keeping warm when outdoors in winter.

(Manuscript received December 18, 2008;
revision accepted for publication July 28, 2009.)