6

Artifact Categorization: The Good, the Bad, and the Ugly

Barbara C. Malt and Steven A. Sloman

1. INTRODUCTION

A sizeable subfield of cognitive psychology is devoted to how humans categorize entities in their world, yet there has been little explicit consideration of what it means to categorize. One straightforward perspective, the one perhaps held at an intuitive level by most laypeople, is that categories are objectively defined, metaphysically real groupings of entities. The process of categorizing is thus a judgment about which grouping any given entity belongs to, a judgment that is objectively right or wrong. Applied to artifacts, this perspective suggests that when people see an object such as a table or bench, they judge to the best of their ability what metaphysically real category of objects this particular object belongs to.

If pressed, though, most cognitive psychologists would probably claim to be agnostic about whether or not there are objectively defined, metaphysically real groupings of artifacts. Relegating such determinations to the realm of philosophy (see, in this volume, Searle, Elder, and Thomasson, all of whom argue for the metaphysical reality of artifact categories), they would suggest that what they mean by artifact 'kinds' are psychological kinds: groupings recognized by humans that might or might not correspond to the kinds that would be identified by philosophers as metaphysically real (see Medin and Ortony 1989). Given this reformulation, the process of artifact categorization is the process of judging which psychologically real, if not metaphysically real, grouping an object belongs to. This framing would likely be subscribed to by most cognitive psychologists studying artifact categorization (if not by most philosophers; e.g. Thomasson, this volume).

To study how an artifact gets categorized, then, one must choose a task or tasks to reveal how a judgment is made about what grouping it belongs to. And in order to make that choice, one must have an idea of what the relevant psychological groupings are; a task that requires judgments about groupings that are not psychologically meaningful or natural would be of little value. An almost universal assumption is that the psychologically real groupings are stable groupings that map directly onto mental representations constituting 'concepts' (see e.g. Keil, and Mahon and Caramazza, this volume) These concepts in turn map onto names and images, and serve induction, planning, and other higher processes. Under this assumption, nouns provide a convenient entry-point into the system—the words *ball* and *doll* delimit psychologically real groupings of objects, are associated with distinct concepts and images in memory, and so on—and so are frequently used to index the categories of interest, but they do not otherwise have special status. (Hence, for instance, in the developmental literature, word and concept learning are often not distinguished; see e.g. Kelemen and Carey, this volume.) Following from the assumption of stable groupings that map directly onto names, concepts, images, and so on, selecting an appropriate task is relatively simple: one can study categorization through any of a number of tasks, such as naming, sorting, property projection, or clustering in memory. All will tap the same groupings.

But closer examination suggests that both of these assumptions are faulty. Any given artifact can participate in more than one grouping. The same rubber ball may, at different times, be grouped with other spherical, bouncy objects, with other things called *ball* (including non-bouncy beanbag balls and non-spherical footballs), with other toys such as dolls and board games, or with other things to take to the playground such as a tricycle and a snack. What the relevant grouping is depends crucially on the situation at hand (Barsalou 1983, 1991; Ross and Murphy 1999; Sloman and Malt 2003). And different tasks may invoke different processes that operate under different constraints. Naming, for instance, is a task that engages language in order for communication to take place, and it must therefore be sensitive to influences such as a language's history and the particular history of a speaker and addressee. In contrast, tasks such as storing information in memory, projecting properties, or planning in order to achieve goals are not about communication; each has its own unique demands (Malt *et al*. 1999; Malt, Sloman, and Gennari 2003*a*).

Below, we discuss the variety of distinct mental activities that people engage in in daily life that can reasonably be considered 'categorization', and we analyze the cognitive demands of each activity. We consider how these activities relate to tasks used in research on categorization. We discuss the nature of the laboratory tasks involved and how they influence the category judgments that are made, and we argue that the tasks used often do not map well onto the activities of daily life that they are meant to shed light on. We review central findings about how people group artifacts that have accumulated and suggest that some of these are useful

with respect to understanding one or more of the mental activities we identify as categorization. We also suggest that some of them are unlikely to contribute usefully, given the discrepancy between the nature of the research tasks used and the nature of the activities of interest. We suggest that given the distinct nature of the activities that involve grouping artifacts, each must be understood on its own terms. To achieve this understanding, researchers must commit explicitly to the type of categorization they are interested in and select methodologies that are appropriate to that type. We argue further that because the term 'categorization' does not carve the space of human endeavors at its joints, no coherent account of artifact categorization is possible, and 'categorization' is not a coherent field of inquiry.

2. CONNECTING OBJECTS TO WORDS

A prominent form of everyday mental activity that involves judging what grouping an artifact belongs to is that of connecting objects with words. In speaking (and writing), people frequently produce names for objects. In almost every utterance, they use nouns, and many of the nouns are intended to refer to artifacts. To produce the utterance, they will usually have an intended referent in mind. They then select a name for the entity: *hat*, *table*, or *ball*, etc. In doing so, they are in effect grouping the object with other objects that have the same name. In comprehension, people hear (or read) artifact names and interpret them by connecting them with objects, real or hypothetical. In some cases—as in hearing 'Hand me that hammer'—potential referents are physically present and the addressee must decide which among them is most likely being called by that name. In other cases, referents are not present—as in discussing needing to buy a hammer—and the addressee must construct in her mind a potential referent. In both cases, the person is, again, in effect grouping the object or objects with other objects that have the same name.

A large proportion of research on artifact categorization uses tasks involving connecting objects with words. The bulk of the research uses a type of task we will call 'name appropriateness judgments'; a smaller portion uses free naming. For each type of task, we first consider its relation to how objects get connected with words in daily life and then review and evaluate findings that have emerged from studies using the task.

2.1. Name Appropriateness Judgments

In most instances of connecting objects to words in ordinary discourse, there are many options on one end or the other. The speaker chooses among many names; the listener chooses among many possible referents. On relatively rare occasions, a more restricted and explicit form of naming choice is called for.

Such judgments are required in response to questions like 'Is that a telephone?' or 'Is that a telephone or a fax machine?' In such situations, one object and only one or two names are under consideration. Situations of this sort arise mainly when someone sees an object with unfamiliar properties and seeks information about the nature of the object by asking about the appropriateness of a particular name.

Despite the rarity of explicit, constrained name appropriateness judgment in the real world, it is the most common type of problem posed to participants in research on artifact categorization. The task differs from most naming choices in daily life in that it combines elements of both production and comprehension. Names are given (as in comprehension) rather than generated (as in production) but the referent is already known to the respondent, unlike in most comprehension situations, and respondents must decide if the name(s) offered correspond to ones they would produce for the object. Further, laboratory name appropriateness judgments have several distinctive characteristics not shared even by the explicit judgments in real world situations. First, the communication goal for the respondent has little to do with achieving reference. In real-world situations, the goal is to illuminate the puzzled inquirer; in laboratory situations, it is merely to satisfy an experimental requirement for a response. Second, because the primary goal is not to achieve reference, discourse context plays relatively little role in determining appropriateness. In real-world situations, context may have some influence even in explicit name appropriateness judgments in that the respondent may consider whether the names offered would suit the questioner's presumed communication needs. In the laboratory, however, there is little basis for inference about communication needs. Finally, response possibilities in the laboratory task are highly limited. In forced choice versions, one of a fixed number of options, often two, must be selected; there is no option to choose neither or more than one, nor to choose a modified version of a name offered (a respondent cannot say 'it's a cell phone' or 'it's a phone-fax' in answer to 'Is it a telephone?'). Likewise, rarely does the respondent have the option of indicating that a name might be acceptable but only marginally so, or acceptable under some circumstances. In one-option versions, the offered name can be rejected but the respondent still cannot offer alternatives or modifications.

In the forced choice laboratory task, then, the participant must consider the names given and decide which of the alternatives is a better name for the object, with the discourse goal and context providing only weak constraints on the operationalization of 'better'. In the one-option version, the participant must consider the name given and judge its acceptability, again with weak or unclear constraints on what should be acceptable. The key decision on the part of the participant is thus what sort of criterion (or criteria) to use in judging the appropriateness of names offered.

2.1.1. Findings

The goal of studies using name appropriateness judgments is usually, though not exclusively, to determine what type of information is most important in artifact categorization or serves as the 'core' of artifact concepts (e.g. Ahn *et al.*'s 2001 response to Strevens 2000; Kemler Nelson, Russell, *et al.* 2000). Among the bases that have been examined are the form (physical features), current function, original function intended by the creator, category membership intended by the creator, and features having a particular causal status with respect to other features.

2.1.1.1. Form vs. Function

The most extensively studied contrast is that of form vs. function. Results of some studies have suggested that form drives naming choices, whereas others have favored function. Some have suggested a developmental trend from form to function, and others have suggested that neither form nor function heavily dominates choices.

In an early study, Gentner (1978) showed children and adults two novel complex objects, named them, and demonstrated their functions. Participants were then shown a test object that had the form of one but the function of the other and asked whether it should be called by the name that had been associated with the same function or the name that had been associated with the same form. Young children and adults both preferred the name associated with the form (though older children tended to make more function-based choices). Rips (1989) presented verbal descriptions of objects created with an intended function associated with one common name (e.g. *lamp*) but an appearance associated with another (e.g. *umbrella*) and asked people to choose between the two names. In contrast to Gentner's results, he found that people preferred the one associated with the intended function.

Several subsequent studies found evidence for variable responses within a single paradigm. Landau, Smith, and Jones (1998; see also 1992) presented novel or familiar objects along with names and, in some conditions, demonstrated the object functions. They then asked children and adults whether objects having either the same function or same shape should be called by the training name. They found that the children tended to accept the name for objects with similar shapes. Adults tended to accept the name for those with similar functions, but they did so more for unfamiliar objects than familiar ones and more when function had been demonstrated during training than when not. Malt and Johnson (1992) used verbal descriptions of objects and varied whether the objects had the physical or functional features normally associated with familiar artifact names. Their participants (adults) judged some objects that had the normal intended function but unusual physical features to be acceptable examples of the name, but not others. At the same time, participants rejected some objects that

had the normal physical features associated with a name but an altered intended function, but they accepted others. Malt and Johnson concluded that neither physical features nor intended function alone fully determine whether people would view an object as an acceptable example of an artifact name category.

Kemler Nelson, Frankenfield, *et al*. (2000) suggested that people may reject a name associated with the stated function if that function is not a plausible explanation for the physical properties displayed. They presented 4-year-olds with novel complex objects and taught them a name for the object. In one condition, the object's function was plausibly connected to its physical features; in the other, the function was something that the object could do but did not appear to have been designed for. Participants chose which of two objects—one similar in function but dissimilar in form, and one similar in form but incapable of performing the function—was another example of the name. The children showed a much stronger bias, though not absolute, to choose objects that preserved the function of the training object when the function had been plausibly connected to the physical features.

2.1.1.2. Current Function vs. Intended Function

Another contrast that has been examined is that of the object's current function vs. original intended function, again with conflicting results. Keil (1989) showed kindergarteners through fourth graders pictures of familiar artifacts (e.g. a coffeepot) and then described alterations that gave the object both the appearance and function associated with a different type of object (e.g. birdfeeder). He asked children to choose between the name associated with the original version and the name associated with the new features. In this case, with both appearance and use altered, the children had a strong tendency to prefer the name associated with the current function and appearance. Matan and Carey (2001) presented 4- and 6-year-olds and adults with ambiguous pictures of objects (pictures in which the object was partially hidden) and told them that the object was made for one purpose but was currently being used for another (for instance, an object made to be used as a watering-can was currently being used as a teapot). Participants were asked to judge whether the object belonged to the first category or the second. In contrast to Keil's result, all participants tended to favor the original intended function, with adults doing so more than children.

2.1.1.3. Intended Category Membership vs. Form and Function

The intended category membership of an object refers to what category the creator of the object had in mind for it. Motivated in part by several free naming studies examining the impact of intended category membership on naming (see below), Chaigneau (reported in Barsalou, Sloman, and Chaigneau 2004) used scenarios describing familiar objects or variations of them to test the relative impact of several factors including the intended category membership on name judgments. In one type of scenario, the object had the usual properties of an

object such as a mop, but it was created accidentally (not intended to be a mop). In another, the object was made by its creator to be a mop but it did not have typical mop features and would not function very effectively as a mop. In others, the object was made by its creator to be a mop but it was used by an agent to perform other actions, or was made by its creator to be a mop but was used to wipe up water only accidentally. Participants judged whether the object was an example of the name associated with target category. Chaigneau found that intended category membership had some influence on judgments but mattered less than the form and use of the object.

2.1.1.4. Causes vs. Effects

Several studies have examined the relative impact on name appropriateness judgments of features that serve as causes for other features vs. those that are effects of other features. Ahn (e.g. Ahn 1998; Ahn and Kim 2000) proposed that causes would be treated as more critical to category membership than effects. Ahn *et al.* (2000) gave participants descriptions of named objects, including artifacts, that specified the causal relations among the features (one feature was the underlying cause of the others; another was an intermediate cause, and the third was the effect of the others). Test objects were missing one feature or another, and participants judged how likely the objects were as examples of the named category. Objects missing the most fundamental cause were judged least likely to be an example of the named category, and objects missing the effect were judged most likely. However, Sloman, Love, and Ahn (1998) argued that effects of feature centrality (on naming as well as other tasks) were not due to causal relations per se but rather to dependency relations in general: a feature is central for naming to the degree that other features depend on it. In a similar vein, Rehder and Hastie (2001) suggested that the critical variable may not be depth in a dependency graph, but rather the number of causal relations that a feature participates in. They presented adult participants with information about categories of artifacts (cars and computers) including attribute values and base rates of attribute values in the categories. In some conditions, participants were also given information about causal relations among the attributes, with the attributes being either causes or effects of other attributes. In transfer trials, they were asked if a described object was a member of the learned category or not. Which attribute was weighted most heavily in the decision depended on the number of relations it was involved in, not on whether it was a cause or an effect per se.

2.1.1.5. Influence of Background Knowledge

Murphy and Medin (1985) suggested that people's background knowledge or naive 'theories' about the world influence how people understand the relation between features of entities and judge their relevance to category membership. A number of studies examining the role of background knowledge in categorization

have used other sorts of tasks or stimuli (e.g. Kaplan and Murphy 2000; Spalding and Murphy 1996), but Lin and Murphy (1997) focused on artifacts in a name appropriateness task. They asked participants to learn names for sets of novel objects, giving participants in different conditions different explanations of how the objects were used. They then asked whether test objects missing one feature of the learned set were examples of the name categories. Participants gave different judgments to the same objects depending on how central the missing feature was to the explanation they had been given. For instance, when training objects were long and narrow with a handle on one end and a loop on the other, a test object with a missing loop was less likely to be judged to belong to the training category when participants had been told that the training objects were for catching animals than when they had been told that they were for spraying pesticides.

2.1.2. Implications of Name Appropriateness Data

By their nature, the strength of laboratory name appropriateness judgments is in revealing what criteria, of those made available by the stimulus construction, a person uses to decide which of two presented names is better for a test object (in the forced choice version) or whether a presented name is justified (in the one-alternative version). Although experiments using name appropriateness judgments have exploited this strength to try to determine what type of information is most important in artifact categorization or serves as the 'core' of artifact concepts, the results suggest that a number of different types of information can influence name judgments of this sort. Different factors gain importance depending on the information made available by stimulus construction and experimental demands. These factors include the form of the object, its original intended function, its current function, its intended category membership, and the structure of the relations among its features. One factor may dominate another in an experiment in which two are selected to be the main sources of variation in the stimuli, but the aggregate results do not point to one factor as the sole or primary basis for judgments in name appropriateness tasks. The body of research using name appropriateness tasks has thus failed in the attempt to identify specific types of information that are most central to artifact categorization or serve as the core of artifact concepts, although it does offer the conclusion that a variety of different factors may influence judgments.

A second implication that emerges from the data is that people do not treat the manipulated features as independent pieces of information. Instead they often try to understand the causal relations of the features to one another, and they consider how this relation relates to that of typical objects associated with a name (Barsalou, Sloman, and Chaigneau 2004). This suggestion emerges from studies on causal relations among features (Ahn *et al.* 2000; Rehder and Hastie 2001) and those on background knowledge (Lin and Murphy 1997), from Kemler Nelson, Frankenfield, *et al.*'s (2000) manipulation of the relation of intended

function to form, and from informal evidence in both Matan and Carey's (2001) and Malt and Johnson's (1992) studies. Matan and Carey noted that adults, and to a lesser extent children, often gave justifications for their decisions that drew not only on the original intended function but on the relation of the functions to possible forms. For instance, a participant might reason that a frisbee can more feasibly be used as a plate than vice versa and so judge that an object that has been used as both is more likely to really be a frisbee. Consistent with Kemler Nelson's suggestion that judgments are influenced by the plausibility of the relation of the physical features to functions, Malt and Johnson noted that their objects with unusual physical features that were most reliably accepted as examples of a name tended to have an interpretation as viable futuristic versions of current objects. Because evidence of this tendency to understand and interpret relations among features emerges from studies that are not designed to make such relations salient as well as from those that are, it appears to be a pervasive feature of how people deal with artifacts rather than one made prominent only by the nature of the experimental tasks.

2.2. Free Naming

As already noted, in daily life, people frequently make judgments about artifact groupings in the process of language production. They generate nouns intended to refer to artifacts in many of their utterances. We call the production situation 'free naming' because the names considered and produced for a given object are not restricted to a small set explicitly provided by an external source. In contrast to name appropriateness judgments—rare in the real world but common in the laboratory—free naming is relatively less studied in categorization research despite its ubiquity in daily life.

Free naming has several characteristics that set it apart from name appropriateness judgments. Memory demands are greater because speakers must retrieve potential names from memory. The set of names stored in memory is extremely large, and the subset activated by an object may be more than just one or two, so the choice set at several points in the production process is potentially much larger than in name appropriateness judgments. In addition, because the speaker may choose to name in whatever way he or she finds most useful, a given head noun can be modified in various ways. Finally, in real world discourse, free naming usually has a specific goal—typically, the goal of achieving reference for some intended audience.[1] The context in such situations can help constrain understanding of the intended referent by an addressee and so may influence what name(s) can be used to successfully refer to it. Free naming within the

FN:1

[1] Naming in ordinary discourse may have additional, usually secondary, goals as well, such as conveying attitude (e.g. in calling a dwelling a *hovel* vs. *palace*). In adult speech directed to young children, teaching names may also be a frequent goal.

laboratory shares most of the characteristics of free naming in daily life except, importantly, that (as for laboratory name appropriateness judgments) the goal for the respondent has less to do with achieving reference than with satisfying an experimental requirement for a response, and the context provides little constraint on what name would be a useful choice. However, some studies create semi-natural communication contexts for naming.

2.2.1. Findings

Evidence about artifact categorization from free naming comes both from observations of naming in the real world and from laboratory situations in which participants are presented with stimuli and asked to name them. The data have been used to address a range of issues about how people name artifacts.

2.2.1.1. Form vs. Function as a Basis for Naming

Laboratory free naming data have been used to address the controversy over whether form or function dominates naming choices for artifacts, and they suggest that neither is clearly dominant. Sloman, Malt, and Fridman (2001) examined the ability of three different versions of similarity—physical, functional, and overall—to account for names produced for three sets of sixty objects: bottles, jars, and other small containers for food; boxes, cartons, and other storage containers; and bowls, plates, and dishes. Participants gave judgments of the similarity of the objects to one another (via either sorting or pairwise similarity ratings), judging either similarity of form, of function, or overall similarity. They were then asked what they would call each object in ordinary conversation. A prototype, nearest neighbor, and weighted sum model were applied to the similarity data to predict the names given. Results showed no clear advantage of one type of similarity over any of the other types in predicting names. A small advantage was seen for physical over functional information, but it was not consistent across models. A subsequent study entered features obtained in a feature-listing task into a Bayesian probability model to try to predict names for the same stimuli. No single feature could predict names fully, but the feature with the greatest predictive power for each stimulus set was a physical, not a functional, feature.

These outcomes are compatible with observations of naming in the real world, which suggest that names are sometimes extended based on form, sometimes on function, and sometimes on a substantial or partial match to both (Malt 1991; Malt *et al.* 1999). Table 6.1 provides examples of objects that appear to share a name based primarily on similarity of form, and Table 6.2 provides examples that appear to be based primarily on similarity of function. Table 6.3 provides examples of objects that appear to share a name based on substantial or partial similarity of both form and function. All examples in the tables were observed within discourse contexts (although recorded without details of the contexts). Note that cases of each are rarely pure; for instance, although the

Table 6.1. Examples of names extended on the basis of form

Name	Typical example	Form-based extension
blanket	bedroom blanket (keeping warm in bed)	picnic blanket (sitting on)
boat	sailboat (transportation)	jail boat (holding convicts)
bowl	soup bowl (holding and eating liquid)	pasta bowl (holding and eating solids)
		sugar bowl (holding and serving granules)
box	shoe box (holding solids)	juice box (holding liquids)
broomstick	kitchen broomstick (sweeping)	witch's broomstick (flying)
chair	kitchen chair (sitting)	electric chair (killing)
fork	from place setting (bringing food to mouth)	carving fork (holding meat)
		fish fork (serving from platter)
		tuning fork (making sound)
gun	pistol	label gun
		glue gun
		staple gun
knife	kitchen knife	frosting knife (spreading)
paper	note paper (writing on)	waxed paper (covering food)
		tissue paper (wrapping gifts)
		fly paper (catching flies)
		toilet paper (wiping)
spoon	soup spoon (scooping liquids)	pasta spoon (lifting pasta from cooking water)
		grapefruit spoon (serrated edge for cutting)
		slotted spoon (draining)
umbrella	rain umbrella (keeping rain off)	beach umbrella (keeping sun off)
wheel	on auto for tires	steering wheel
		Ferris wheel
		spinning wheel

Note: Functions are provided in parentheses for some examples to highlight contrasts or when function might be unfamiliar.

primary function of an electric chair is quite distinct from the primary function of a kitchen chair, the similarity of form does afford a shared component of function in that one sits in both. Likewise, although shared function seems to be the primary reason that manual razors and electric razors share a name, their shared function leads them to be not entirely dissimilar in form. In general, form and function will tend to be correlated, which may be why Sloman *et al.* found that neither was distinctly superior at predicting the names that artifacts in their sample received. The examples in the tables are categorized according to whether the primary link appears to be one or the other, or whether both appear to contribute about equally to motivating a shared name.

Table 6.2. Examples of names extended on the basis of function

Name	Typical example	Function-based extension
auto	Honda Accord	Model T Ford
can opener	manual opener	electric opener
chair	kitchen chair	beanbag chair
computer	Pentium 4 PC	ENIAC (room-sized; vacuum tubes)
corkscrew	with metal spiral	with two prongs
		with propellant cartridge
fan	electric box fan	Japanese paper fan
drill	power drill	dentist drill
		hydraulic drill
key	traditional door key	hotel doorkey (plastic card, magnetic strip)
		electronic car door key (remote control)
razor	manual razor	electric razor

Note: Forms are provided in parentheses for some examples to highlight contrasts or when form might be unfamiliar.

2.2.1.2. Variability in Naming Patterns Across Languages

Languages may have different naming patterns for the same sets of artifacts. Kronenfeld, Armstrong, and Wilmoth (1985) found that speakers of English, Hebrew, and Japanese partitioned a set of eleven ordinary drinking vessels by name in different ways. For example, English-speakers asked to name the objects called a paper drinking vessel and one for drinking tea both *cup*, but Israelis labeled them with different names. Japanese participants used three names in partitioning the objects, but they were partitioned by only two different names in English and in Hebrew. Malt *et al.* (1999) asked speakers of American English, Mandarin Chinese, and Argentinean Spanish to name sixty common containers and found substantial differences in the naming patterns across the three languages along with similarities. Malt, Sloman, and Gennari (2003*b*) examined the relation among the linguistic categories of the three languages for the sixty containers in more detail and found a complex pattern. Some of the categories shared prototypes across the three languages but others did not; some cases of nesting occurred (the categories of one language were contained within those of another); and some cases of cross-cutting were found (pairs of objects were put into a single category by one language but into different categories by another language). These divergent patterns have consequences for second-language learners: Malt and Sloman (2003) found that non-native speakers of English failed to match native naming patterns for these same sixty containers and for sixty examples of housewares, and some discrepancies persisted even for non-natives having many years of immersion in an English-speaking environment.

2.2.1.3. Variability in Name Choices within a Language

Even for a given language, artifacts often have more than one acceptable name, both across and within people. Malt *et al.*'s (1999) data showed that although

Table 6.3. Examples of names extended on the basis of shared form and function or partial overlap of both form and function

Name	Typical example	Multi-factor extension
bed	in bedroom (sleeping)	sofa bed (sitting and sleeping)
bottle	Coke bottle (holding and drinking liquid)	aspirin bottle (holding solids)
		spray cleaner bottle (holding and spraying liquid; spray top)
box	shoe box	litter box (no lid)
		bread box (metal; curved)
		lunch box (metal; domed top)
		plastic animal-shaped juice box (for liquids; has straw)
camera	film-based	digital camera
chair	kitchen chair	dentist chair (for dental procedures, not just sitting)
		rocking chair (for rocking, not just sitting)
fan	electric box fan	ceiling fan
knife	kitchen knife	butter knife (cutting and spreading)
		paring knife (peeling)
		butcher knife (chopping)
		cheese knife (forked tip for spearing)
ladder	extension ladder (for climbing up to reach)	step ladder (for climbing up to reach)
		rope emergency ladder (for climbing down to escape)
oven	electric oven	Dutch oven
		microwave oven
		toaster oven
phone	touch-tone phone	cell phone
		1920s dial phone
table	kitchen table	drafting table (slanted; for drawing)

Note: Forms or functions are provided in parentheses for some examples to highlight contrasts or when they might be unfamiliar.

consensus among speakers of a given language was very high for some familiar containers (e.g. a glass juice bottle was called *bottle* by all participants), for other members of the set names varied considerably (e.g. a gallon container of milk was called *jug* by some, *container* by some, and *bottle* by others). In a pre-test for a matching task involving artifacts (see below), Malt and Sloman (2004), using free naming followed by name verification, readily identified a number of familiar objects that had two commonly used names. (For example, one object was called a trashcan, a wastebasket, and a trash container, and another was called both a booklet and a pamphlet; individual participants verified that all names were acceptable.)

2.2.1.4. Previous Exposure Effects

Given the existence of more than one potential name for an artifact, the question arises of what determines choices among them on any given occasion. One factor appears to be the creation of an implicit agreement between participants in a conversation about what name will be used to refer to an object (e.g. Brennan and Clark 1996; Clark and Wilkes-Gibbs 1986). Brennan and Clark (1996) had pairs of participants carry out a matching task in which one participant arranged the pictures in an order described by the other. In the course of the task, participants established tacit agreements about how to refer to the pictures, taking into account the level of specificity needed to discriminate the objects from one another. Brennan and Clark found that these agreements, once established, influenced naming in later discussions even when the context of those discussions would have allowed reference to be achieved with simpler expressions. For instance, if a picture set contained two types of shoes, participants gave them names such as *sneaker* and *high heel*; they continued to use such names later for picture sets in which only one type was present and *shoe* would have been a sufficient label.

Another factor influencing the choice is what names have been recently retrieved from memory and used, regardless of their relation to any previous agreement with a conversational partner. Sloman, Harrison, and Malt (2002) asked participants to name artifact stimuli created by morphing two familiar artifacts (e.g. a pen and a marker). Participants had previously been exposed to other objects that were named with one label or the other. Names given for the target stimuli were influenced by the previous exposures. Malt and Sloman (2004) had pairs of participants discuss artifacts pre-tested as having two acceptable names (e.g. *trashcan* and *wastebasket*) in carrying out a matching task. One participant was a confederate who introduced one of the two names for each object. In subsequent trials, the naive participant performed trials of the same task with another naive participant. Names used with the new partner were influenced by which name was initially used by the confederate, suggesting that a bias toward a particular name, once established, carries over (at least in the short term) beyond conversation with the original partner.

2.2.1.5. Sensitivity to Intended Category Membership

Bloom (1996) suggested that naming is sensitive to the intended category membership of an object's creator: People name artifacts in accordance with the category that they think the creator intended for it. Bloom and Markson (1998) asked 3- and 4-year-old children to draw pictures of a lollipop, a balloon, the experimenter, and themselves, and later asked them to say what the drawing was a picture of. The pictures of lollipops and balloons were generally not distinguishable from one another, nor were pictures of the experimenter and the child, but children of both ages tended to name pictures in accordance with their original intention in producing the drawing. Gelman and Bloom (2000)

described objects as either intentionally or accidentally created; for instance, a newspaper was folded into the shape of a hat either by a person or by being run over by a car. Children and adults were asked what the objects were. All age groups gave an object name (e.g. *hat*) more often than a material name (e.g. *newspaper*) when the origin was intentional and vice versa when it was accidental.

2.2.2. Implications of Free Naming Data

The free naming data that exist have been collected under conditions that capture many of the demands and constraints of naming in daily life. A shortcoming of much of the research using free naming tasks is that it does not provide natural discourse contexts and goals for naming, and so outcomes cannot reflect their potential influence. Some exceptions do exist, however (e.g. Brennan and Clark 1996; Malt and Sloman 2004), providing insight into these influences. The aggregate data provide a number of important pieces of information about naming.

The free naming data indicate that there is no unique grouping of artifacts by name, either between languages or within languages. Patterns of naming for the same artifacts differ across languages, and the same artifacts can receive different names from different speakers within a language, and even from a single speaker on different occasions. The variability observed here does not involve labels at different levels of abstraction (e.g. *table* vs. *furniture*), nor does it involve fundamentally different types of groupings such as those invoked in the service of momentary goals (e.g. things to take out of the house in a fire; Barsalou 1983, 1991) vs. those that are taxonomic (e.g. *car*). Rather, the same objects may be grouped by name differently using basic level, taxonomic labels as a function of the individual speaker, the language he or she is speaking, and the circumstances of the utterance.

Some of the variables that affect name choice can be thought of as short-term influences in that they influence the choice made by a speaker on a particular occasion from among those names available in his or her language. The effects of recent retrieval episodes on what is retrieved from memory for a subsequent stimulus is one such factor (Sloman, Harrison, and Malt 2002; Malt and Sloman 2004). The others involve adjusting to the conversational context in ways that are under greater speaker control. These include taking into account the name agreed upon with a conversational partner (Brennan and Clark 1996), and taking into account the presumed intention of the creator (Bloom and Markson 1998; Gelman and Bloom 2000) even when no name has as yet been explicitly offered and accepted.

In contrast, the variability in naming patterns across languages appears to be the consequence of longer-term factors that establish what names are available for objects within a language and what the preferred assignment of names to objects within a domain is. Malt *et al.* (1999, 2003*a*, 2003*b*) suggest that the name for a given object in any particular language is influenced by what names happened

to exist in that language at some earlier time and so were available for extending to new objects; what objects happened to exist in the culture at some earlier time and either formed a similarity cluster that was given a name, or extended outward from a cluster and caused a name to be extended to less similar objects by chaining (Brugman 1983; Lakoff 1987; Taylor 1995); what names happened to be bestowed on objects by a manufacturer either from within that culture or from outside for marketing purposes; and what domains were of particular interest to a culture at some point in its history and so required finer linguistic differentiation of the conceptual space. The consequences of each for a given language will vary as a function of the linguistic and cultural history involved, and so languages will diverge to some extent in their naming patterns even when showing commonalities driven by shared perception of the stimulus space.

From the longer-term perspective, perhaps it should not be surprising that languages differ in their naming patterns for the same set of artifacts, nor that names can be shared across objects based on similarity of form, function, or both. Within the domain of artifacts, possibly more than within any other domain, new variations on existing entities are created on a frequent basis. Each new variant is not likely to be given a unique name unless the intention of the maker is to isolate it from its predecessors. Instead, existing words will be extended to cover new cases. Often the variations for each new object, either in form or in function, are relatively minor and so the object retains many of the properties of its predecessor. For users of the language, there is little to impede comprehension if the word is extended to cover such variants.[2] However, the cumulative result of these small steps of extension can be that a word is associated with a set of objects that vary considerably in form, function, or both. For reasons suggested above, languages may follow different paths of extension, with the result that they accumulate different sets of objects sharing a name.

The observed diversity in how objects are grouped by name might seem to work against the usefulness of names for conveying information about objects (that is, allowing the addressee to identify a physically present or hypothetical object and make appropriate inferences about the properties of the object referred to). However, other characteristics of the free naming task in real world

[2] Petroski (1993) provides a fascinating example of how the knife and fork evolved in Western culture. In the 1500s table knives were narrow and had a pointed tip, and were used not only for cutting meat but also for spearing food and conveying it to the mouth. Pointed tips later gave way to blunt tips for safety reasons. At that time, forks had only two tines and were used primarily for holding meat steady while it was being cut. Foods that could not be easily speared by the two-tined fork, such as peas, were conveyed to the mouth by piling them on knives, and knives developed a wide blade that was bulbous at the tip to provide a better surface. Later, forks evolved to have three or four tines and became used as the primary means of bringing food to the mouth. The wide, bulbous blade of knives then reverted back to a narrower, straighter style. Each step in this sequence involves a relatively small change, but both the forms and functions of knives and forks changed over the course of the evolution. The names 'knife' and 'fork' (with some variation in spelling) were used in English for these objects throughout this time-period (*Oxford English Dictionary* 1989).

situations may make this diversity unproblematic. The discourse context can greatly constrain interpretation of a noun. If a gallon container is just out of reach of a person wanting to fill it with water, it matters little if the person asks for the *bottle*, *jug*, or *container* to be passed, because any of the names will be sufficient to achieve reference. In addition, the open-endedness of the free naming task allows speakers to use noun-noun or adjective-noun combinations to clarify referents and direct inductions. Someone hearing a noun such as *bottle* or *chair* will look for (if potential referents are physically present) or imagine (if not) different objects and induce a partially different set of properties depending on the modifiers attached. Hearing *electric chair* will lead a person to assume different properties than hearing *kitchen chair*, as will hearing *baby bottle* vs. *aspirin bottle*. Brennan and Clark's (1996) data indicate that speakers adjust the specificity of names as needed to disambiguate intended referents from other potential referents in the discourse context. The open-ended nature of the task also makes it possible for speakers to take into account what name they think their conversational partner intends to be applied to an object or what name they have previous established with the partner for referring to the object, which contribute to the likelihood of successfully achieving reference.

In addition to the specific information provided by the data about how naming choices are made, these data make evident the extent to which naming is an activity that is embedded in a linguistic and communicative context and that reflects the demands and possibilities made available by this context. We suggest that it will be impossible to understand name choices for artifacts without considering communicative issues such as what a particular addressee can understand or will understand most readily (reflecting speaker—addressee naming history and the availability of discourse context to constrain understanding, among other things), and linguistic issues such as the historical linguistic forces that shape the vocabulary available to a speaker of a given language and the language's conventions for applying that vocabulary and the availability of modifying phrases to accomplish goals of naming. Research that has a goal of understanding how people choose names for artifacts but that eliminates most or all of these influences from the judgment process is not likely to substantially advance knowledge about naming.

3. CONCEPTUAL (NON-LINGUISTIC) GROUPING

So far we have been considering forms of categorization that involve connecting objects to words in order to communicate. However, people also group artifacts in other sorts of situations where the primary goal is not to communicate about an object. We now consider a range of other mental activities occurring in daily life that can reasonably be considered 'categorization'. For each, we discuss the laboratory tasks using artifacts that have the greatest resemblance to that

particular activity, and we review findings from those tasks and implications of the findings.

3.1. Object Recognition: Connecting Objects with Stored Knowledge

Perhaps the most frequent form of artifact grouping in daily life is making sense of objects encountered by connecting them with stored knowledge about objects. When someone walks into an office, for instance, and sees an object with a flat wooden surface, four legs, and drawers, he or she recognizes the relation of the object to previously experienced objects (and/or abstractions across them) having similar properties. Doing so allows the person to interpret the visual input in terms of the past experience, yielding an understanding, for instance, that the surface of the perceived object will be rigid and afford writing, that the drawers may contain paper-clips and stationery, etc. This recognition process goes on almost continuously, as people move about the world encountering and making sense of objects. In a familiar environment, much of the time the process will be one of connecting familiar objects with stored knowledge of the same objects (as in recognizing one's own desk, rug, armchair). Sometimes the process will be a matter of connecting a novel object with knowledge of very similar objects (or abstractions)—as for a not-previously-encountered chair of traditional design—and occasionally it will be a matter of connecting an object with more novel features to knowledge about objects that may be less similar as for example, in an initial encounter with a chair in the shape of a hand or a chair made of rope that is hung from a ceiling.

Notably, this recognition process does not require engaging language. A person can appreciate the relation of a visual stimulus to stored knowledge without needing to retrieve a name for the object. This is amply evident from the fact that such recognition occurs in many situations where a name is not available for an object: an infant who has no word for a bottle nevertheless interprets the bottle offered to it by virtue of previously experienced bottles; a chimpanzee interprets a tree never encountered before by virtue of previously experienced trees (see Hauser and Santos, this volume, for discussion of artifact concepts held by non-linguistic animals); an adult human who is introduced to a novel object will recognize a second object of the same type as such without knowing a name for either. Indeed, even familiar, everyday objects are sometimes appreciated without being named. Many people discriminate several species of birds in their neighborhood without knowing names for them, and Malt *et al.* (1999) found that some participants had no well-established name for certain common objects, such as a plastic container of baby powder with a shaker top. When pressed for a name, participants resorted to phrases such as *a thing* of baby powder. To the extent that encountering a desk may tend to bring to mind the word *desk*, this word retrieval likely occurs as a *result* of having made contact with stored non-linguistic information associated with the word, rather

than the word retrieval preceding and enabling contact with the non-linguistic information. Recognizing artifacts in the environment is thus an activity most centrally having to do, not with language, but with the processing of visual input and with memory retrieval and comparison processes.

A second characteristic of this form of grouping is that there are no boundaries to the groupings and hence no discrete categories to which the object belongs or does not (Malt *et al*. 1999; Sloman and Malt 2003). Any given object simply has a resemblance to each previously encountered object to a greater or lesser degree. For instance, a newly encountered wooden object for sitting on that has a round seat, four legs, and a low back will make contact with stored knowledge about various other forms of seating that it resembles, and some of those may have taller backs and arms while some have neither. The former may be called *chair* and the latter *stool*, if naming is required, but coming to understand the new object itself requires no choice between the two. A gradient of relevance to the novel object may exist, but retrieval of stored information across this gradient is not constrained by category boundaries.

3.1.1. Findings

3.1.1.1. Name Appropriateness Judgments

When psychologists studying artifact categorization talk about a person categorizing an object as a chair, or table, or desk, and so on, they often seem to have in mind the non-linguistic process of connecting an object in the world to stored knowledge of similar objects. However, perhaps because communicating about such processes almost inevitably involves using names—we talk about someone categorizing an object 'as a chair' or 'as a desk'—researchers typically have not discriminated between the non-linguistic process of interest and the process of naming. As a result, studies that may be designed to shed light on non-linguistic categorization often use the choice of a name as the dependent measure. They ask participants whether they would call the object a chair or a stool, a cup or a glass, and so on, and in doing so they conflate the process of selecting a name with the process of object recognition. We have already considered the large literature using name appropriateness judgments from the perspective of how naming choices are made, and have argued that understanding naming as it occurs in daily life requires studying naming as part of a linguistic and communicative system. Because this literature has, as we argued, tended to present naming choices in an impoverished context that does not fully engage this system, might it, in fact, usefully shed light on object recognition instead? Unfortunately, the use of names as a response measure and the failure to discriminate naming from the non-linguistic recognition process has led to a focus on issues that appear to be more relevant to naming than to object recognition. For instance, debates about whether form vs. function or current function vs. original intended function are more influential in determining name choice seem to have little relevance to understanding how an encountered object

makes contact with knowledge about objects stored in memory. Furthermore, because the task requirement is to make a choice between discrete categories whereas object recognition requires no such choice, the task itself seems to have little bearing on the recognition process. The primary useful finding from this literature may be the notion we highlighted earlier, that people do not treat object features as independent pieces of information but rather attempt to understand the causal relations among them. This causal analysis is presumably relevant to appreciating the nature of an object non-linguistically, not only to choosing a name.

3.1.1.2. Similarity Judgments

A frequent assumption in the literature on artifact categorization is that the task that most directly reflects the representational structure tapped by the object recognition process is similarity judgments. If perceived objects evoke stored knowledge by virtue of their features, then similarity judgments have the potential to reveal the basis on which a gradient of relevance for retrieval of that stored knowledge operates. (Similarity itself, of course, may also play a role in the groupings generated by other mental activities, as we will discuss later.)

When people recognize the relation of objects in their environment to stored knowledge of similar objects, they often do so without conscious awareness, and the end result is simply an understanding of the nature of that object. In laboratory similarity judgments, however, participants usually make deliberate comparisons of presented objects and provide an explicit judgment about their likeness. In some cases, similarity judgments are collected by presenting pairs of objects (usually in the form of pictures) and asking for a numerical similarity rating. In other cases, participants sort objects (again, usually pictured) into piles, and a measure of the similarity between each possible pair of objects is derived from the number of participants who sort them into the same pile. The extent to which making such judgments conscious, requiring an overt response, and having the judgments be entirely among physically present objects (instead of comparing an object at hand to stored knowledge of objects) alters the pattern of grouping indicated is unclear.

Two salient results have been obtained from judgments of artifact similarity. First, perceived similarity is remarkably constant across cultures, despite the variable naming patterns discussed earlier. When Kronenfeld *et al.* (1985) asked the same native speakers of Hebrew, English, and Japanese who produced divergent naming patterns for drinking vessels to sort them into groups according to their similarity, they found that the correlations among interpoint distances in multi-dimensional scaling solutions of the similarity matrices ranged from. 81 to. 89. Malt *et al.* (1999) had the same speakers of English, Spanish, and Chinese who produced divergent naming patterns for sixty common containers sort the pictures into groups according to their overall similarity. They found correlations among the similarity matrices ranging from .91 to .94. The high

degree of consensus across cultures suggests that the featural analysis of artifacts is universal in nature (perhaps based on a universal causal analysis relating form to function, as suggested by the studies discussed earlier) as are comparison processes and will yield a shared understanding of the nature of objects.

Second, despite the consistency of similarity judgments (in parallel neutral contexts) across cultures, perception of similarity for any given population is not fixed across all contexts but rather is influenced by the nature of the comparison at hand. For instance, a given object A may be judged more similar to object B when in the context of object C than when in the context of object D. Medin, Goldstone, and Gentner (1993) found a variety of context effects on similarity judgments for stimulus materials that included some artifacts (see also Barsalou 1983, and Ross and Murphy 1999). Shiftings of perceived similarity based on the comparison objects themselves may be infrequent in actual object recognition, where the perceived object is compared to a relatively stable base of stored knowledge rather than to a small and deliberately varied comparison set, but it does suggest that the broader context in which the perceived object occurs may alter the gradient of relevance.

3.1.1.3. 'Object Perception' Tasks

A large literature examining how objects are perceived exists outside of that traditionally considered to be categorization research. This literature has addressed a range of topics including how objects are isolated from a complex visual scene, whether recognition is orientation dependent, what the relative roles of parts, outlines, and shading are, the effects of expertise, whether processing occurs at different scales, and whether different brain systems subserve perception of different classes of stimuli such as faces vs. objects. In doing so it has used tasks including word—picture matching, naming, old/new judgments, and familiarity judgments. Reviewing this literature is beyond the scope of this paper, but we note that it does directly examine some of the processes that are central to object recognition. From that perspective, this literature is better focused on issues that are genuinely about recognition. However, as the list of issues above suggests, and as the phrase sometimes used to label this field—'visual object recognition'—implies, the work has focused primarily on lower processes that are part of the initial visual processing that must take place in order for a percept to make contact with stored knowledge. As such it does not directly illuminate how stored knowledge about the properties of objects (beyond their visual features) is brought to bear on understanding the nature of a perceived object.

3.2. Induction

Another form of grouping is induction, in which people use information about one (or more) object(s) to infer properties of others. Induction is not a process that laypeople would be inclined to call 'categorization'. However, it depends

crucially on judgments of whether and why entities are alike and as such can be thought of as a form of categorization.

One frequent situation in which induction occurs in daily life is in the recognition process. When someone recognizes the relation of an object with a flat wooden surface, four legs, and drawers to stored knowledge of objects sharing some or all of these properties, that recognition provides the basis for making inferences about properties not readily apparent. When visual input is interpreted in terms of past experience with the same object (e.g. viewing your own desk), and the perceived object is projected to have properties that are not deducible directly from the visual features (e.g. understanding that the surface is rigid and will afford writing, and that the drawers can open and will contain stationery), an inference has been drawn. Likewise, when there is no stored knowledge corresponding directly to the perceived object but it is interpreted by reference to stored knowledge of similar objects (e.g. walking into an unfamiliar room and viewing a desk never seen before), properties of the object are projected from experience with similar known objects. Induction that occurs during the recognition process is typically rapid and non-conscious, and the particular inductions generated are presumably a function of the memory comparison and retrieval processes that take place during recognition.

Induction also occurs in everyday situations in which the inductive process is slower and more deliberate. On a relatively infrequent basis, when recognizing objects, a person may encounter an object with quite novel features and engage in conscious consideration about the nature of unseen features, given knowledge of other objects. For instance, a person viewing an unfamiliar kitchen gadget that fits over the neck of a wine bottle and has a needle may infer that it is made for removing corks, based on familiarity with other forms of corkscrew. More often, the conscious projection of properties occurs within a learning context. When a person learns from external sources or discovers through direct experience with an object that an object has a property previously unknown to him or her, the person may conclude that certain other objects have that same property. For instance, if a person learns that the ink in her ballpoint pen has a certain chemical composition, she may infer that the ink in other similar objects (e.g. other brands of ballpoint pens) probably has the same composition. In such cases, the induction is from a single object to others, but inductions may also be from a group of objects to others. For instance, if a person learns that all waterproof inks have a certain substance in them, she may infer that the ink in her ballpoint pen does. In either case, the person considers what other objects might sensibly be considered to have the same property as the known one(s), given the nature of the known object(s) and other possible ones.

Names can be useful cues to the appropriate projection of properties, since objects that are labeled with the same name tend to share at least some properties in common. Much of the research on induction in the developmental literature has focused on how children make use of category names to guide inductions

(e.g. Gelman and Markman 1986; Davidson and Gelman 1990; Gelman and O'Reilly 1988). Nevertheless, as with recognition, induction does not inherently involve language. One can project a property from one object (or a set of objects) to others based on beliefs about the nature of the objects, without reference to their names. Indeed, names are not always reliable cues to properties, especially for artifacts. The fact that baby bottles are made to be unbreakable does not imply that Coke bottles are. (However, it may imply that toddler 'sippy' cups are.) The essential demand in inductive situations is to draw appropriate conclusions about properties of objects that are not readily apparent, regardless of how they are named.

Although induction does not inherently involve language, names are, in fact, present in many inductive situations that involve learning because information transmittal often takes place via language. Rather than discovering new facts about an object through direct observation, people are often told of such facts. In such cases, an added demand to the task of projecting properties is the retrieval from memory of knowledge associated with the words. For instance, if someone is told that 'All carpentry tools are subject to a stiff tariff', in order for her to consider what objects might be subject to a stiff tariff, the phrase 'carpentry tools' must activate information in memory and cause retrieval of knowledge of particular objects or sets of objects. The objects retrieved may or may not be the full set of objects that could reasonably be called 'carpentry tools', and the particular knowledge about the objects that is activated may or may not be the same as the knowledge brought to mind when actually seeing or using such objects. Thus inductive situations that engage language add elements of memory and lexical access to the task.

3.2.1. Findings

3.2.1.1. What is the Probability that Object X has Property Y?

A common paradigm in research on induction provides information to participants about the properties of one or more objects and asks them to decide whether another object would have that property, or what the probability is that it would have that property. In most research with adults, the objects are presented by means of words, as just discussed.

Sloman (1998) used this paradigm to investigate whether people would use class-inclusion relations in projecting properties. For instance, given information about a property of electronic equipment, and given agreement that stereos are electronic equipment, will participants agree that the property must be true of stereos? Sloman found, for artifacts as well as natural and social kinds, that people did not consistently follow class inclusion relations in their judgments. Instead, they agreed more often when the objects in question were typical (e.g. stereos) than when they were atypical (e.g. kitchen appliances) of the larger grouping. Sloman found that the effect occurred even when people verified that the named objects were electronic equipment shortly before responding to the induction

questions. Although the paradigm in general raises the question of whether the effects obtained are driven by the limitations of memory retrieval in response to the linguistic stimulus, given that this effect occurred even after a recent probe for the information, this particular effect seems less likely to be due to memory retrieval problems upon hearing the phrase 'electronic equipment' than to a reasoning process in which people assess featural (similarity) relations between the set of objects in question and the set encompassed by the superordinate name.

Gelman (1988) used a variant of this paradigm that included pictures to investigate whether pre-schoolers and second-graders were sensitive to differences in the projectibility of properties across different groupings and property types. A child might be shown a picture of a rabbit and told that the rabbit likes to eat alfalfa. The child would then be shown other entities and asked if they thought the others also had the property. The child responded to four test items that varied in their relation to the standard in each case—for instance, another picture of the same rabbit (same name, same appearance), a different-colored rabbit (same name but different appearance), a dog (different basic-level name but same superordinate name—*animal*), and a telephone (different basic level and superordinate name). Half of the standards were artifacts (the remainder were natural kinds). Children showed an almost linear induction gradient across the four test item types, agreeing to the inference less often as degree of relatedness decreased. They also agreed to the property inferences somewhat more often overall for the natural kinds than for the artifacts, but properties concerning function were projected more for artifacts and those concerning substance more for natural kinds. This study provides an important illustration of the fact that induction patterns are not simply a reflection of naming patterns. Children did not consider all and only objects with the same basic-level name to share a property. Rather, children considered the nature of the relation of each entity to the standard and they projected properties based on this relation. They also considered the nature of the property in question along with the nature of the objects in judging whether the property should be projected to the test items. (The same conclusion is suggested by Mandler and McDonough's (1996, 1998*b*) work on induction in children under 2 years old, using an object manipulation paradigm; see also Mandler, this volume.)

3.2.1.2. Forced Choice Judgments

An induction paradigm used in many developmental studies and in some studies with adults presents a new fact about an object and then presents alternatives, of which one (or more) has one type of relation to the first object and the other(s) have a different type of relation to it. Participants are asked which is more likely to share the property. This paradigm has a similarity to the forced choice naming tasks discussed earlier, in that participants have no option to indicate that both choices might support the specified inference, or that neither does, or that other possibilities not given might be better than either choice offered. These studies

can therefore indicate which of two specific options is preferred but do not indicate whether both choices might be acceptable, or what the most likely object to project the property to given free choice would be. Davidson and Gelman (1990), following work by Gelman and Markman (1986) that used only natural kinds, showed children unfamiliar objects, half of which were artifacts, and test objects that either shared a name with the standard or were perceptually similar. They found that the value of names in promoting inferences for children was integrally tied to the conceptual relatedness of the objects associated with the name: if the objects did not share substantial properties in common, children did not prefer inferences based on a shared name. Thus the power of names in these tasks appears to reflect the non-linguistic understanding of relations among objects. Children do not draw inferences based on shared names per se but rather on an assumption that objects sharing multiple observable properties will share unseen properties as well. Farrar, Raney, and Boyer (1992) found similar results using a paradigm more similar to that of Gelman (1988).

Ross and Murphy (1999) used a forced choice task to investigate whether food groupings of different types supported different types of inferences. The new properties they presented were either biochemical or situational, and the choices were between an object with a taxonomic relation to the first or a script-based relation. Thus, for instance, in the biochemical condition they asked: 'Suppose that an enzyme, metascal,• has been found in bagels in the country Quain. What food is more likely to contain metascal: cracker or egg?' where 'cracker' is the taxonomic choice and 'egg' is the script-based choice. The situational property was that the object is used in an annual initiation ceremony. For the enzyme property, choices were primarily taxonomic, and for the ceremony, they were mostly script-based. Consistent with Gelman's (1988) finding contrasting artifacts and natural kinds, Ross and Murphy suggested that different types of groupings support different types of inferences.

• Q1

In sum, although induction research using artifacts is limited in quantity, three central points emerge from it. First, people consider the similarity among objects in judging whether a property that holds true of some is likely to be shared by another. Second, they also engage in more sophisticated reasoning that takes into consideration both the nature of the property and the nature of the objects involved. Finally, the patterns of property projection indicate that induction is not constrained by the linguistic categories associated with objects. Instead, it is based on knowledge of the shared properties of objects and beliefs about whether they will share additional, unseen properties. Names are useful as guides to induction only to the extent that they are indicators of such shared properties.

A much larger literature on induction exists that is focused on natural kinds rather than artifacts (for a review, see Sloman and Lagnado 2005). Some of the central findings in this literature concern the role of similarity in driving inferences. For instance, given the premise that robins have sesamoid bones,

people find the conclusion that sparrows have sesamoid bones more plausible than that ostriches have sesamoid bones, and they also judge robins and sparrows to be more similar than robins and ostriches. This literature also makes clear that when people reason about familiar objects and properties, they also engage in causal reasoning by attempting to explain the relations between categories and predicates. For instance, when told that a particular type of tree has a disease and asked if another tree will have that disease, people having high familiarity with the forest ecosystem may make a judgment based not on the overall extent of shared properties between the trees but rather on what the likely mechanism of disease transmission would be and whether the mechanism is likely to operate among the trees in question (see Medin *et al.* 2002). The findings discussed above suggest that people engage in a parallel analysis for artifacts.

3.3. Planning and Problem-Solving

People also frequently create groupings of artifacts in planning and problem-solving in daily life. Some of the groupings are formed in service of temporary goals (Barsalou 1983, 1991). When planning a trip to the beach, people may retrieve from memory knowledge of objects to take to the beach (a towel, a book, sunscreen, a drink); when considering a baby gift for a friend, they may retrieve possible purchases (a receiving blanket, a silver spoon, clothing, a mobile). Other groupings are formed in the service of enduring goals. In order to meet the goal of recycling properly, people may construct a mental grouping of all the types of objects they should not throw in the trash but should save for the recycling bin. In order to eat, people may maintain mental groupings of foods that are appropriate for breakfast, for lunch, and for dinner (Ross and Murphy 1999). Still other groupings may be formed for recurring as opposed to ongoing goals. For instance, people often need objects to contain substances or items. In each episode of need, they will retrieve knowledge of a set of objects that would suit the materials to be contained: round glass containers with wide mouths to hold the firefly their child caught; round and squarish plastic containers with lids to hold their dinner leftovers. Retrieval of appropriate knowledge for enduring and recurring goals may become highly practised and stabilized.

As with recognition and induction, knowing the names of the objects is usually incidental to formulating the thought or solving the problem. Retrieval may activate names associated with objects, and activating the names may be critical to communicating the thought or action plan to someone else, but the central demand of the task is to choose objects that will effectively serve the goal regardless of object names.

3.3.1. Putting Objects Together that 'Belong Together'

A laboratory task that has been little used in connection with artifacts is asking participants to sort objects according to what objects belong together (rather

than on their similarity per se). This task presumably taps the more stable, practised groupings used in planning and problem-solving. Its infrequent use is perhaps because researchers are inclined to think that they already know what stable groupings people use—those labeled by common nouns such as *table* and *ball*. However, the set of objects labeled by a particular noun need not be the grouping used in any given goal-directed task, as our discussion indicates. Some stable groupings may have no single word name, and some single words may label a diverse set of objects only some of which are relevant in the service of any particular goal.

Ross and Murphy (1999) provide one of the few cases of this type of sorting task using artifacts. They asked people to sort foods such as cereal, hamburger, and milk. They found that when instructed to, participants were able to readily form both groupings based on the composition of the foods (e.g. putting dairy products together) and groupings based on the role the foods play in daily life (e.g. putting breakfast foods such as cereal, eggs, and bacon together). When allowed to sort without direction, they spontaneously produced some groupings of each type. This outcome underscores the fact that the same objects can participate in multiple groupings having different bases and indicates that people may maintain such cross-cutting groupings in memory rather than having a strictly taxonomic organization of knowledge.

The sorting task as Ross and Murphy implemented it does differ crucially from natural situations in that there was no task context establishing a particular goal to be served by the groupings. Presumably if specific goals were induced, participants would tend to use whichever type of grouping best served the goal at hand.

3.4. Organizing Novel Information

Finally, people group objects when they are confronted with an array of novel objects and try to make sense of them by constructing groupings of entities sharing important properties. For adults, this situation arises relatively rarely but may occur when beginning to learn about a new domain. For instance, a person taking a job in an electronics company may need to learn about a large number of different electronic devices, varying in form and function, that she has never encountered before. This person seeks to understand the domain by constructing groupings of the devices sharing one or more properties. For children, being confronted with an array of novel objects in a previously unfamiliar domain may arise more frequently as they explore the world around them.

As with the other tasks we have discussed in this section, knowing or using names is not inherent in the process of organizing novel information. One can appreciate the shared properties among objects and form groupings without having names for them. Indeed, in cases of true discovery—for instance, a scientist exploring a distant land and encountering an array of insects or plants

never before known to science—conceptual organization of the new entities must precede generating labels for the groupings. As with induction, though, in many cases language is a medium of input into the formation of groupings. The adult learner who encounters a new domain may not be left to create the groupings entirely on his or her own. He or she receives some input about appropriate groupings from other adults in the form of names for the objects and sometimes explanations of their properties or reasons for being grouped together. For young children, as well, the process of grouping the objects based on shared properties is often guided in part by input from adults in the form of labels and explanations.

3.4.1. 'Category Construction'—Creating Groupings from Arrays of Novel Objects

A number of studies in the general categorization literature have used a task that has been labeled 'category construction'. In this task, participants are given an array of novel objects that they are asked to place into groups in whatever way seems natural. The groupings are examined to determine on what basis they have been formed. This paradigm closely resembles real-world situations in which a child or adult is exposed to a new domain containing many unfamiliar entities and seeks to organize knowledge of them without external input. It differs from those real-world situations in lacking input from others in the form of labels or explanations about useful groupings, and in lacking a task goal to suggest what sorts of groupings might be most useful.

The bulk of studies using this paradigm have not used artifacts as stimuli. Because of the need to provide novel entities to group, many have used highly artificial entities as stimuli such as geometric shapes. In a few cases, the stimuli can be liberally interpreted as artifact-like. Ahn and Medin (1992), using starfish-like shapes that did not necessarily appear animate, found a strong tendency for people to base their groupings primarily on a single dimension (and add in anomalous exemplars at the end). Regehr and Brooks (1995) tested sorting of various shapes under a variety of conditions and found a bias toward one-dimensional sorts but that more sorting based on multiple dimensions occurred when participants did not see all the stimuli at once, and when they did see them all at once but alternated which group they were adding to rather than building one in its entirety and then the other. Wattenmaker (1992) used verbal stimuli consisting of four features, including some describing artifacts, and found a heightened tendency for people to rely on a single dimension when they were making groups based on memorized rather than physically present stimuli. These results are compatible with previous studies using non-artifact stimuli (e.g. Medin, Wattenmaker, and Hampson 1987) demonstrating a bias toward one-dimensional sorts, but they also indicate that the attentional demands under which participants construct the groupings influence whether they attend to multiple dimensions or just one.

Other studies suggest that sorting strategies are influenced by the knowledge base that is engaged by stimuli. Lassaline and Murphy (1996), using stimuli that included artifacts, found that participants were more likely to create multi-dimensional sorts when their attention was first drawn to feature correlations within the stimulus set by answering induction questions (e.g. 'If a vehicle has bench seats, what kind of top does it have?'). Kaplan and Murphy (1999) found that multi-dimensional sorting increased when stimuli could be related to familiar 'themes' (e.g. arctic vehicle vs. tropical vehicle). Ahn (1999) provided background information about stimuli by manipulating whether the features of the prototypes were described as all having the same cause, or all as causing the same effect, or whether the features are related in a causal chain (one causes the second which causes the next, and so on.) Participants were more likely to create multi-dimensional sorts when they had causal knowledge that allowed them to construct explanations of feature variability (i.e. the common cause and common effect conditions; the causal chain provides no explanation for feature variability because each feature predicts a fixed other feature). Category construction tasks thus show that people have a tendency to group objects along a single dimension in laboratory tasks but that they may group based on multiple similarities under some circumstances. They also show that when provided with richer stimuli and information about causal relations among their features, people make use of this information to create groupings that overlap on multiple dimensions.

3.5. Implications of Data on Conceptual (Non-Linguistic) Categorization

The aggregate data from the laboratory tasks that are not centrally about connecting objects to words, along with our analysis of task demands, highlights several facts about non-linguistic groupings. First, the perception of similarity is an important influence in determining the groupings. The empirical evidence explicitly implicates judgments of similarity in the projection of properties. But our analysis of task demands suggests that it must also be involved in other forms of non-linguistic grouping, influencing what stored knowledge is retrieved in interpreting objects encountered in the environment, what objects will be grouped together in the course of planning and problem-solving, and what objects will be grouped together in making sense of novel domains. Second, analysis of the causal relations among properties of artifacts occurs in these non-linguistic tasks (as it appears to in naming tasks), and, further, in tasks that involve conscious reasoning (as in the more deliberative instances of induction), causal reasoning about the relation of properties of one object to properties of another may take place. Third, despite these common components, the groupings produced in each case are not necessarily the same. This fact derives, in part, directly from the varying demands of each type of mental activity. The previously experienced objects that will be most useful in interpreting an encountered object will not

necessarily be the same as the set of objects to which one might project a particular property of the encountered object, nor will it be the same as the set of objects it might be grouped with to serve some temporary goal, or to organize knowledge of the domain if the entire domain is novel. The difference in groupings also derives, more indirectly, from the flexibility of perception of similarity given such differing task demands. The data show that under parallel task demands, people from different cultures speaking different languages nevertheless share perception of commonalities among artifacts, presumably because they all understand the object properties and their causal relations to one another in the same way. However, the data also demonstrate that what properties are most relevant to the judgment of commonalities varies depending on the task demands. As our analysis makes clear, different mental activities that involve grouping artifacts non-linguistically do make distinct demands.

The relative scarcity of data using artifacts in such tasks, though, and the rather loose connection of the tasks to instances of categorization in daily life, indicate that there is much left to learn about the nature of artifact groupings that are formed in the course of the mental activities that can be considered conceptual categorization.

4. GENERAL DISCUSSION

In our analysis so far, we have identified five types of mental activities occurring in daily life that can be thought of as categorization and considered the requirements of these activities. We have reviewed laboratory tasks that have been used to shed light on categorization, discussed the relation of the requirements of these tasks to those of the activities of daily life, and reviewed findings from those tasks. We now consider what has been learned about the nature of artifact categorization from the research reviewed, and we draw out the larger implications from our analyses for what the study of categorization should look like.

4.1. The Good

A number of important findings have accumulated from categorization research. We have already summarized the central findings about different forms of artifact categorization and highlighted those we consider most valid. In the case of linguistic categorization, these include that the form of an object, its original intended function, its current function, and its intended category membership may all influence judgments of name appropriateness. In addition, people actively seek to understand the causal relations among features of an object and they consider how the relations relate to that of typical objects associated with a name when they judge name appropriateness. When naming freely, short-term variables influence choices including what names have recently been retrieved

from memory and what name for an object has been agreed upon with a conversational partner or appears to be the name the object's creator has in mind. Longer-term historical variables also contribute to naming choices; these establish what names are available for objects and what the preferred pattern of assignment of names to objects within a domain is.

In the case of non-linguistic categorization, several findings also stand out. To the extent that explicit similarity judgments reflect the kind of similarity gradients that must operate in recognition, it can be concluded that what will be perceived as similar in the recognition process is stable across cultures when the context of judgment is similar, but that context may shift what objects are seen as most similar to each other. In induction, the groupings reflected in patterns of property projection are influenced by the similarity among objects and by reasoning processes that take into account the nature of the properties and objects involved; names are a useful guide to appropriate projections but only to the extent that they are understood as indicators of shared properties among the objects. In planning and problem-solving, the same sets of objects may be grouped in different ways depending on the goals and contexts of the situation. In organizing novel information, people draw on their understanding of the causal relations among objects to create groupings of objects that overlap on multiple dimensions.

From these findings, several more general conclusions also emerge. Each has already been noted in the context of non-linguistic grouping tasks; considering them in conjunction with the findings on linguistic categorization makes clear the generality of the phenomena. One general theme is that similarity is a crucial contributor to most or all forms of groupings. We have discussed its involvement in all the variants of non-linguistic grouping. It appears to play a role in determining how perceived objects are connected with stored knowledge, how properties are projected, in how novel information is organized, and in what groupings are constructed to serve goals in planning and problem-solving. In the case of naming, although cross-linguistic differences in naming patterns are striking, we also noted a degree of cross-linguistic consistency that is presumably driven by a shared perception of similarity among objects.

A second general theme is that the analysis of causal relations among the properties of artifacts occurs spontaneously and is a fundamental part of people's appreciation of the nature of the artifacts. Causal analysis influences people's choice of names and also their grouping behavior in non-linguistic tasks such as property projection and grouping novel stimuli.

Finally, a third theme emerges from contrasting the nature of the groupings that arise in each of the five types of mental activities, and that is that the particular groupings formed vary substantially according to the activity involved. For instance, linguistic categories do not map directly onto the groupings involved in induction. Groupings formed in the service of a particular goal may have little

to do with the way that objects from the same domain would be grouped when making sense of them as a novel array.

Can these findings, then, in some way be integrated to begin to provide a truly general and complete theory of artifact categorization? We first review what we consider weaknesses of the research on artifact categorization, and then ask whether a more careful approach will solve the problems and permit development of a unified theory of categorization.

4.2. The Bad

Although some useful findings have resulted from the categorization research we have reviewed, we believe that other arenas of investigation have been less fruitful. In particular, findings in some cases have accumulated from paradigms that do not map well onto the nature of the mental activities in which grouping takes place in daily life. If the demands of the research task are substantially different from those of the grouping task in daily life, results from the task are unlikely to be useful in shedding light on the mental activities of interest.

Results from the highly studied forced choice version of name appropriateness judgments are a major case in point. As we have pointed out, explicit name appropriateness judgments are rare in the real world, and so a task that requires this sort of judgment is not likely to capture the nature of most naming. In particular, because of the demands of this task, it is likely that any factor manipulated in controlled contrasts will show an effect. The paradigm requires that participants always choose between two options somehow. Although random responding could occur, a cooperative participant responding to the experimental demand for a judgment is likely to seek some systematic basis for choosing. If an experimenter were to present named objects having particular colors and sizes, and then present test objects having the color associated with one label and the size associated with the other, participants would most likely select either color or size as the basis for their responses rather than respond at random. If a consistent choice emerged across participants, that would suggest that participants see the chosen dimension as more important to naming than the other. However, it would not indicate that the dimension chosen is the sole basis for naming decisions outside of the experimental context. Further, it might not even indicate that one dimension is given more weight outside of the forced choice contrast. For instance, participants may treat function as more important than form in making the explicit judgments because they feel they can justify a judgment based on function more easily than one based on form. In natural naming situations, though, choices may be pushed in one direction or the other by factors outside of the speaker's conscious awareness. Indeed, the free naming observations we have discussed suggest that the preference for function over form that has tended to show up in forced choice data is not mirrored by natural name extension patterns.

Of course, as we pointed out earlier, it is not clear that researchers using name appropriateness judgments are always interested in naming in particular. They do not necessarily talk of 'naming' but rather simply 'categorization'. This raises the question of whether name appropriateness judgments may, instead, be useful as a task for tapping the groupings that are demonstrated in the course of understanding objects encountered in the environment by reference to stored knowledge. As we argued, the task seems poorly suited for studying this form of grouping activity because it requires a choice between names, whereas recognition does not involve choosing between discrete categories. In addition, because the task measure is a name choice, whereas recognition does not inherently involve language, it can at best be a somewhat second-order reflection of the processes involved.

But the tendency of researchers who use name appropriateness judgments to often talk about 'categorization' rather than 'naming' as the issue of interest highlights the somewhat schizophrenic nature of much research on artifact categorization. On the one hand, methodologically, the research treats names as if naming is what categorization is about—that is, the dependent measure is what name an object is judged to have. On the other hand, the discussion that introduces the issues and tasks used is typically not about how objects are named. It is divorced from any considerations of the communicative function of naming, how reference is achieved, how languages evolve patterns of name extension, or the like. Likewise, the tasks themselves rarely are designed to engage naming in any natural sort of discourse context. The primary interest often appears to be something about how objects are grouped non-linguistically: how objects are put into 'categories.' Name choices clearly are attractive as a dependent measure because they provide a tractable overt behavioral response. Nevertheless, we propose that if the issue of interest is not naming, other sorts of measures are necessary. In particular, if the question at hand has to do with how people understand an encountered object by reference to stored knowledge, then the response measures need to be ones that are revealing of how visual input is processed and how memory retrieval and comparison processes operate.

We noted earlier that forced choice judgments about induction—for instance, asking participants whether they prefer to generalize a property to an object that shares a name with the standard or that shares perceptual features with it—have some of the limitations of forced choice naming judgments in that they require the participant to choose one option, when generalizing to neither or both might be reasonable given free choice. Like the naming paradigm, they can reveal the importance only of those options that the experimenter presents, and so run the risk of making especially prominent those factors that current theorizing specifies as of interest but that may not be most central to inductive judgments in general.

Also potentially problematic in the induction arena is the paradigm widely used with natural kinds as well as in some studies involving artifacts in which

participants are asked to judge the validity of conclusions given a premise that is introduced verbally. If reasoning as it takes place when mediated by language is the issue of interest, then the paradigm may be well suited to the issue. If the question of central interest is, however, how people understand the nature of some objects by drawing on their knowledge of the properties of others, then introducing information about objects verbally adds demands—in particular, lexical access and retrieving information associated with words from memory—that may alter responses. Indeed, the type of inference about properties that takes place in the rapid, non-conscious appreciation of objects when encountering them in the environment is surely not mediated by language and operates in a distinctly different time-scale than the processes that are engaged when language is involved.

Other tasks that seem of limited value include sorting tasks intended to illuminate the groupings used in planning or problem-solving that do not specify goals or contexts for the sorting, and category construction tasks that do not allow the participant to engage any background knowledge or construct a causal understanding of relations of stimulus properties. Free naming tasks that involve no discourse context or communicative goal will also be less revealing of naming in daily life than those that do.

Finally, there is a large literature that we have not reviewed in this paper in which people are asked to learn to divide abstract patterns into groups and then are tested to see which group they believe a test stimulus belongs with. Because of the need to create novel stimuli to be learned, artifacts are not generally used in such research. However, this literature has the goal of understanding how categorization in general, presumably including artifact categorization, takes place, and so we note our concern with it here. Research in this tradition is not intended to be about naming; category designators are usually arbitrary labels such as '1' and '2'. The main concern in this literature appears to be with how a newly encountered stimulus is associated with stored knowledge; in our terms, with the object recognition process. We have argued that in daily life this process does not involve bounded categories, nor does it involve making any choices between groupings. The relevance of this paradigm to the process of interest thus seems quite limited, in parallel with our comments on the relevance of forced choice name responses for experiments using artifact stimuli.

Given that we have argued that a number of types of task commonly used to study categorization may not be good choices, one might ask whether there is a 'right' task to use. The answer to this must be 'no'. An essential requirement for a good choice of task is that it engages the processes that are normally engaged in the real-world activity it is meant to shed light on. As we have argued repeatedly by now, there is no single real-world activity that constitutes categorization. As a result, there can be no single right task for studying categorization. Many tasks may be appropriate, but their appropriateness can only be judged against a clear specification of what form of categorization is of interest. Is it how

people connect objects with words, and if so, in the case of production or comprehension? Or is it how people understand the nature of objects in the world as they encounter them by making contact with stored knowledge and by drawing inferences? If the interest is in inference, is it in those inferences that may be drawn rapidly and non-consciously in the process of recognition, or in those that are drawn in more deliberative situations, perhaps when a new fact has been learned about one object? Or is the activity of interest what groupings people construct or retrieve from memory in the course of planning and problem-solving? Or about how people deal with organizing knowledge of a set of novel objects?

Little progress will be made if researchers treat 'categorization' as a topic of inquiry that needs no further differentiation or analysis. Selecting appropriate methodologies requires identifying what the mental activity of interest is and carefully analyzing the demands and constraints of that activity. Our suggestion is not intended to serve to champion ecological validity in experimentation for its own sake. Indeed, we believe that the purpose of an experiment is not to mimic the world but to explain it. Our point is that in order to explain, one must know what one wants to explain, and one must use methodology that will be revealing of the thing to be explained. In the absence of any common referent in everyday behavior, the class of 'categorization' tasks is itself artifactual.

4.3. The Ugly

Thus far, we have been discussing the problems of choosing appropriate tasks to reveal the nature of the categorization activities of interest. But we believe that there is also a more profound problem for categorization research than the need for researchers to make explicit the form of categorization that they are interested in and select research methodologies accordingly. The deeper problem is revealed in the diversity of the groupings resulting from the different types of mental activity, reflecting the different demands and constraints of each type of activity. The problem is that the term 'categorization' does not carve the space of human endeavors at its joints. The sets of mental activities that can reasonably be called 'categorization' are diverse and operate in different ways to accomplish different ends. Indeed, in the traditional terms by which cognitive processes are organized in textbooks and allocated to journals, some of the mental activities we have discussed would be labeled 'higher cognitive processes' or 'thinking' (the conscious induction situations; planning and problem-solving; organizing information in novel domains), some would fall under 'language' (naming), and others would be considered 'lower cognitive processes' and given labels such as 'pattern recognition' or 'object perception' (what we have called 'object recognition' and the associated rapid, automatic inductive processes). These traditional labeling distinctions reflect the differing sets of issues that must be addressed in order to understand the processes involved. Thus we believe

that there will not be any useful or coherent account that covers the various forms of mental activity involving grouping artifacts and sets them apart from other components of higher and lower cognitive processes. In other words, 'categorization' is not a coherent field of inquiry. We propose that it will be more fruitful to abandon the goal of having a general theory of artifact categorization and instead focus on understanding, each on its own terms, what people do when they produce and understand names for artifacts, when they recognize artifacts and draw inferences about their properties, when they draw on stored knowledge of artifacts in the service of plans and goals, and when they organize their understanding of a novel domain.

A second consideration also argues against the coherence of 'categorization research' as a field of inquiry. Underlying the notion of categorization research is the assumption that not only stable but bounded categories exist. Categorization researchers talk about objects being members of categories and people putting an object into a category. Such talk may be a result of the fact that both in ordinary discourse and scientific discourse about objects, communication about objects requires using names—we refer to an object as 'a chair' or 'a table'—and names impose discrete structure on continuous conceptual space. But we have argued that much of what researchers are interested in when they talk about 'categorization' does not actually involve bounded groupings. As we have discussed, understanding objects in the world by reference to stored knowledge is what researchers often seem to mean when they talk about categorization. Our analysis of what is involved in making sense of an object by connecting it to stored knowledge indicates there is no need to place the object in any discrete category in doing so. Likewise, there are no discrete boundaries that limit the projection of properties nor the set of objects that may be usefully grouped to meet a goal. Although we have at points followed tradition and talked about the mental activities that yield groupings of artifacts as artifact 'categorization', once the notion of fixed categories, and indeed of bounded categories at all (except in the case of naming), is discarded, then it is not clear that it makes sense to talk about artifact 'categorization'.

Of course, some cognitive processes surely exist that are common to two or more categorization tasks and that map onto some coherent psychological system. The data we have reviewed suggest that a comparison operator that carries out similarity judgments is one, and a causal reasoning system is another. Additional candidates include a perceptual integrator, a decision-making system, and memory storage and retrieval processes. All of these processes are worthy of study. But none of them is unique to grouping activities. Studying them across the different kinds of activities that they participate in may be more useful for the development of cognitive theory than focusing on their operation within only the set of activities that involve grouping.

4.4. Are there such Things as Artifact Kinds?

Finally, we consider two issues that are raised by our arguments about the nature of artifact categorization. First is the question of whether there are such things as artifact kinds.

At the outset of this paper, we noted that psychologists talk of artifacts as coming in 'kinds', where kinds are taken to be stable, psychologically real groupings. We have argued that not only are the psychologically meaningful groupings not stable, but there are no bounded groupings in conceptual space except by virtue of names associated with objects. These arguments suggest that the notion of psychologically real artifact kinds is not viable (Sloman and Malt 2003). The intuition that artifacts do come in kinds, though, remains strong. For instance, in ordinary discourse, it is common to speak of two objects as being the same kind of thing. And such talk is not idle chatter; it usefully serves a speaker's goal of highlighting the existence of commonalities among the objects. Thus some account is needed of why the intuition of kindhood is part of ordinary discourse and what it means for objects to be the same kind of thing in these cases. Notably, however, providing such an account need not require a notion of artifact kinds that are stable and clearly bounded. In ordinary discourse, although it is often relevant to speak of objects as being the same kind of thing, whether any two objects actually are considered the same kind of thing may depend on the context and goal of the particular discourse taking place. A wooden kitchen chair and a beanbag chair may be called the same kind of thing in some circumstances (for example, if being contrasted with tables); on the other hand, the kitchen chair and the beanbag chair may be thought of as different kinds of things in other circumstances (for example, when a wooden desk chair and a wooden rocking chair are also salient in the context). If the goal is to find firewood, the kitchen chair and a broom-handle may be the same kind of thing. Under scrutiny, then, even the lay notion of artifact kindhood is flexible and task-dependent. Thus it may make perfect sense to talk about two objects or a set of objects being the same kind of thing, even while it is impossible to define stable groupings of artifacts that can be identified as the members of an artifact kind.

4.5. Given the Considerations about the Nature of Categorization, What is a Concept?

A final issue is what our arguments suggest for thinking about concepts. Much of the goal of research on artifact categorization is to reveal the nature of artifact concepts. That is, the grouping behavior that can be overtly observed is of interest because it is taken as an indication of the contents of a mental representation underlying it. Finding out what type of knowledge determines

grouping choices is assumed to directly illuminate the nature of stable, coherent packets of knowledge that reside in memory.

Two aspects of our analysis create problems for this assumption. First, and most important, the fact that there is no unique segregation of artifacts into groups raises a puzzle. If there is no unique segregation of artifacts into groups, what grouping or potential grouping corresponds to something that should be called the concept?

Second, in the case of categorization as naming, factors appear to influence the groupings that names pick out that are not directly represented in individual language users' knowledge base. For instance, speakers of English label a stuffed seat for one person with the same name they use for a wooden seat for one person (calling them both 'chair') but speakers of Mandarin Chinese label the stuffed seat for one person with the same name they use for a stuffed seat for multiple people (although English speakers would call the latter 'sofa'), and the source of the difference may be longer-term historical factors rather than anything having to do with understanding of the properties of the objects by speakers of the two languages. Within a single language, the notion that uses of an artifact name may be extended in various directions on different dimensions also suggests that the knowledge associated with groupings picked out by names does not necessarily form a coherent packet (as in the case of a beanbag chair that is called 'chair' by virtue of a functional relation to kitchen chairs and an electric chair that is called 'chair' more on the basis of its form). The common notion in the literature that artifact concepts are packets of knowledge that map directly onto names, thus, in particular, does not appear to be a useful one.

One way to think about what artifact concepts are, rather than taking them to be stable, pre-packaged sets of knowledge that map onto names, is to consider them to be flexible and situation-dependent (e.g. Barsalou 1987; Barsalou and Medin 1986). That is, each time mental activities take place that result in a grouping of artifacts, one could say that a concept has been formed. This approach captures well the idea that the groupings formed in planning and problem-solving vary from occasion to occasion but constitute coherent packets of knowledge. However, it is less satisfying when thinking about the groupings picked out by names, which we have just suggested may not correspond to coherent sets of knowledge. And conversely, the groupings formed in the recognition or induction process may be coherent but seem not to fit the pre-theoretical notion of 'concept' that implies a packet of knowledge that can be consciously appreciated. It also violates the general intuition that a 'concept' should be something stable and resident in long-term memory.

An alternative is to more explicitly identify concepts with the knowledge retrieved when asked about the knowledge associated with the word. This version corresponds to the lay use of the term and captures the researchers' inclination to associate concepts with words. Thus, for instance, if a person is asked what her concept of chairs is, she will retrieve knowledge associated with the word 'chair',

most likely having to do with four legs and a back and arms and being for sitting on. We may consider this knowledge the 'concept' of a chair. This knowledge likely reflects the properties most frequently associated with the word 'chair' and so constitutes a prototype associated with the word. Critically, it does not capture the full set of knowledge that generates the use of the word 'chair' across the full range of potential discourse contexts. Nor does it capture or delimit the range of non-linguistic groupings that may be formed from knowledge of objects that happen to be called 'chair' under other task demands. Thus under this definition of a concept, it is important to remember that 'categorization' is not just about 'concepts', and that an account of 'concepts' will not be derived from studying the range of mental activities we have suggested comprise categorization. By definition, though, it will capture packets of knowledge that are activated in the course of interpreting language and so have some psychological reality. However, it must be noted that the information retrieved by a word may vary according to the discourse context (e.g. Anderson and Ortony 1975; Barsalou 1982; Brauer *et al*. 2003) and so even then, there is no single stable concept associated with the word. And so it may be inevitable that a useful notion of concepts for theorizing about cognitive processes requires discarding the hope of identifying stable packets of knowledge and embracing the construction of concepts in the context of a task.

4.6. Summary and Conclusion

We have discussed the variety of distinct mental activities that people engage in in daily life that can reasonably be considered 'categorization'. We have analyzed the cognitive demands of each activity and considered how these activities relate to tasks used in research on categorization. We have argued that the tasks used often do not map well onto the activities of daily life that they are meant to shed light on. As a result, although some of the existing findings from research on categorization are useful for understanding one or more of the mental activities we identify as categorization, others are unlikely to contribute usefully. We suggest that given the distinct nature of the activities that involve grouping artifacts, each must be understood on its own terms. Further, because the term 'categorization' does not carve the space of human endeavors at its joints, we suggest that no coherent account of artifact categorization is possible, and 'categorization' is not a coherent field of inquiry.[3]

FN:3

[3] We thank Debby Kemler Nelson, Art Markman, and Greg Murphy for helpful comments on an earlier version of this chapter.

Queries in Chapter 6

Q1. Please clarify if the term 'metascal' should be retained as such or if it should be changed to 'metascale' throughout.