

(in press) Language and Cognitive Processes

Artifacts Are Not Ascribed Essences, Nor Are They Treated as Belonging to Kinds

Steven A. Sloman

Brown University

Barbara C. Malt

Lehigh University

Please address correspondence to:

Steven Sloman

Cognitive and Linguistic Sciences

Brown University, Box 1978

Providence, RI 02912

email: Steven.Sloman@brown.edu

Running head: Essentialism, kinds, and artifacts

Abstract

We evaluate three theories of categorization in the domain of artifacts. Two theories are versions of psychological essentialism; they posit that artifact categorization is a matter of judging membership in a kind by appealing to a belief about the true, underlying nature of the object. The first version holds that the essence can be identified with the intended function of objects. The second holds that the essence can be identified with the creator's intended kind membership. The third theory is called "minimalism" (Strevens, 2001a). It states that judgments of kind membership are based on beliefs about causal laws, not beliefs about essences. We conclude that each theory makes unnecessary assumptions in explaining how people make everyday classifications and inductions with artifacts. Essentialist theories go wrong in assuming that the belief that artifacts have essences is critical to categorization. All theories go wrong in assuming that artifacts are treated as if they belong to stable, fixed kinds. Theories of artifact categorization must contend with the fact that artifact categories are not stable, but rather depend on the categorization task at hand.

Psychological essentialism is the hypothesis that object categorization is a matter of assigning kind membership on the basis of a belief about the true, underlying nature of the object. Most of the discussion of psychological essentialism has concerned judgments about naturally occurring entities and their classification into natural kinds. Strevens (2001a) and Rips (2001) both make convincing cases against an essentialist view of everyday categorization for naturally occurring entities. Strevens argues instead for a minimalist view. The minimalist view assumes that categorization is a matter of judging kind membership but that these judgments are based on beliefs about causal laws, not beliefs about true, underlying natures. In this paper, we consider three theories, two essentialist ones and Strevens' minimalist view, and examine their application to artifact categories. Each theory must contend with the fact that artifact categories are not stable, but rather depend on the categorization task at hand.

Essentialist views of artifacts: I. Intended function as essence

Essentialism started life as a theory of word meaning (Kripke, 1972; Putnam, 1975). The original idea was that an individual language speaker using a natural kind term is expressing the term's meaning by appealing to a linguistic convention and not to a mental representation. The arbiter of word usage is an expert (hypothetical or not) who acts as an authority on whether an object is appropriately called by a word. A fundamental requirement of this view of word meaning is that some underlying property or properties that constitute the essence of an individual entity determine the appropriate name for that entity (Gelman & Hirschfeld, 1999). These properties are not necessarily known by an ordinary (lay) speaker and are not even necessarily knowable by such a speaker. Sometimes they are not even fully known by the expert (who may revise his or her beliefs

about the essence as scientific knowledge grows). The minimum criterion is only that properties are assumed that an authority could in principle inspect to evaluate whether an object is appropriately labelled by the kind term.

Psychological essentialism, as introduced by Medin and Ortony (1989), concerns not words, but concepts. It is intended as a theory of how people judge an object's kind. Although it does not make a claim about the existence of metaphysical essences, it shares with original essentialism the ideas that people believe entities have essences and that these beliefs provide the basis for their judgments of an object's kind. As such, it makes two assumptions: first, that people treat objects as belonging to stable kinds; second, that people determine kind by appealing to a shared essence. Our discussion concerns both these issues. Like all discussions of psychological essentialism, our focus is psychological, not metaphysical. We are not concerned with whether kinds in the world truly have essences, but with whether beliefs about essences are causally relevant to everyday behavior. We examine whether beliefs about essences and kinds are causally relevant to judgments about objects' names, properties, and relations to other objects, and not whether those beliefs correspond to metaphysical reality.

According to psychological essentialists (Gelman & Hirschfeld, 1999; Keil, 1995; Medin & Ortony, 1989), people evaluate an object's kind on the basis of their beliefs about the causes of the object's observable properties. They treat the most basic causes, those on which all others depend, as the object's essence. Is it possible to characterize beliefs about such basic causes in the domain of artifacts — human-made entities? If not, doubt is raised about the psychological reality of both essences and, correspondingly, kinds in the domain of artifacts.

One possibility is that artifacts judged to belong to a particular kind all serve a particular function. Something is considered to be a pen if it's used for permanent writing, something is considered to be a boat if it's used for travelling on water. The function would determine the observable properties of the object. Pens are long and thin to fit rigidly in the hand and they contain ink to write with. Boats have a shape that allows them to float while moving forward on water. Therefore, the function of an artifact might be treated as that object's essence, and objects that share that essence will be judged to belong to the same kind. Note, though, that an artifact's function is not actually the most basic cause of its observable properties; in fact, function is usually the effect of an object's observable properties when used by an agent in a certain environment (Barsalou, Sloman, & Chaigneau, in press). The shape of the boat causes it to be able to travel on water. The most basic cause is not its function per se, but its intended function, something in the mind of a user or creator that explains why the object was constructed as it is. So, one essentialist view is that, by virtue of its causal centrality, the intended function of an artifact is treated as its essence (Ahn, Kalish, Gelman, Medin, Luhmann, Atran, Coley, & Shafto, 2001; Keil, 1989; Rips, 1989).

However, the intended function view of artifact essences suffers from several difficulties:

1. Some groups of objects don't have obvious intended functions, and those that do may not be distinguished from one another by their function.

For some objects, like computers or duct tape, it is hard to say what the intended function is. At best one can state only a very general function, such as "to assist in work" or "to hold things together," functions that do not separate these objects from those that

would be called by other names. Even when more detailed functions can be given, such functions may not effectively separate members of contrasting categories (Malt & Johnson, 1992). The intended function of a boat may be to transport people and goods across the surface of the water, but so is the intended function of a raft.

2. An object's intended function is dependent on the particular goals of an agent at a particular time, and on its particular history of use.

Objects can have many intended functions (Barsalou et al., in press). One intended function is that of the artifact's maker (assuming the maker has only one function in mind). But sometimes objects have multiple, independent creators each with a different function in mind, like when a spittoon is turned into an ashtray. The creator of the spittoon has a different intended function for the object than the creator of the ashtray. Furthermore, sometimes the same sort of object is created by different makers for different purposes. For example, broomsticks are usually created with one intended function (to sweep the floor), but sometimes they are created with another (in Harry Potter novels, to fly).

These examples pose a problem for the view that intended function provides the basis for a unique judgment of kind membership for an object. They do not by themselves rule out the possibility that objects can be considered to belong to multiple kinds. Perhaps each intended function of an object places the object in a distinct category. However, as the number of categories that an object is judged to belong to increases, it becomes less plausible that people believe an object has a true, underlying nature that determines its kind membership.

3. When people are asked to judge an object's kind on the basis of either intended function or physical features, physical features are sometimes given more weight. Malt and Johnson (1992) found substantial use of physical features in decisions about artificial stimuli modeled closely on real artifacts (see also Hampton, 1995; Landau, Smith, & Jones, 1998). Some novel objects having the intended function of a familiar category but divergent physical features were judged not to belong to the category, and some having a novel intended function but normal physical features of the category were judged to belong. This pattern is also seen in everyday life, such as in the case of a pizza cutter, which is distinguished from a knife on the basis of divergent form despite sharing the function of knives, and a frosting knife, which is called a knife on the basis of similar form despite divergent function.

Studies examining how much weight people give to functional versus physical information when naming real objects have found corresponding results. Sloman, Malt, and Fridman (2001) tried to predict the names given to common containers (bottles, jars, jugs, boxes, cartons, etc.) and kitchenware (dishes, plates, and bowls) using similarity judgments and three formal classification models: a prototype model, a nearest neighbor model, and a weighted sum model that combined name and similarity information across exemplars. Predictions were made using each of three types of similarity judgment: functional, physical, or overall. No single type of similarity was consistently better able than any other to account for the category names that people assigned the objects regardless of the classification model used. We also examined the degree to which the names could be accounted for by specific features. Two sets of featural descriptions of the objects were fed into a Bayesian classification model to try to predict the names.

Physical features were consistently better than functional features at accounting for names, although not always significantly. Together, these results indicate that function is not given precedence over form in determining what people call common containers.

4. What is the relevant category? Dissociations between naming and similarity judgment imply that beliefs about essences associated with conceptual groupings are not consistent with linguistic categories.

Philosophical essentialism identifies essences with kinds. Change the essence under discussion and the kind changes (by definition); choose a different kind and a different essence is automatically relevant. Such a definition is not susceptible to empirical argument and is not at issue here. Psychological essentialism, however, frames the relation in terms of a judgment process: Beliefs about essences are causally relevant to judgments of category membership; people use their knowledge of essential properties to help them pick out members of a category. Testing such a theory requires two operational definitions: i. a sufficiently well-formulated definition of essence (e.g., intended function), and ii. an independent means to decide what people consider to be in the category (the set of actual or hypothetical objects that they take to constitute a kind). Only with both in hand is it possible to test the psychological essentialist hypothesis by seeing if i. and ii. correspond. But what should serve as operational definition ii., what determines a category extension?

A common assumption, pervasive in arguments for psychological essentialism (e.g., Gelman & Hirschfeld, 1999), is that the category comprises all those objects that are given a particular label. Chairs are those objects called "chair." This simple solution fails though because different languages partition the space of artifacts in different ways.

Kronenfeld, Armstrong, and Wilmoth (1985) had speakers of English, Hebrew, and Japanese name 11 drinking vessels and found that the languages grouped the objects by name differently. For instance, the Americans gave the same name to a paper drinking vessel and one for drinking tea (“cup”), but the Israelis gave them two different names. In a larger-scale study, Malt, Sloman, Gennari, Shi, and Wang (1999) compared the names for 60 common containers given by speakers of American English, Mandarin Chinese, and Argentinean Spanish, and found that the composition of the categories differed across the three languages. For instance, the 16 objects named “bottle” in English were spread across seven different linguistic categories in Spanish. The Chinese category that contained the 19 objects called “jar” in English also included 13 objects called “bottle” in English and eight called “container,” although others called “bottle” or “container” appeared in different Chinese categories. Artifacts do not seem to be grouped into universal linguistic categories; languages categorize them in their own, idiosyncratic ways. Whatever knowledge English speakers draw on in grouping these objects by name, it is not the same as that used by Chinese or Spanish speakers.

Perhaps speakers of different languages just appeal to different essences. This would be fine if the issue at hand were only word use. The linguistic conventions in different languages apparently are different and people could use different criteria in forming beliefs about an object's essence and thereby if the object warranted a particular name. But psychological essentialism isn't most directly concerned with word use; it is a theory about concepts. And according to the theory, the essence is what people believe to be the fundamental cause of the object's observable properties. Speakers of different languages surely don't assume different fundamental causes. The causal laws governing a container

of soda are bound to be understood in the same way by speakers of English, Chinese, and Spanish. Indeed, similarity judgments of these speakers suggest that they are. Malt et al. (1999) found that, even though the speakers had made different linguistic judgments about the objects, they made almost identical similarity judgments. When asked to sort the objects into piles according to their physical, functional, or overall similarity, the differences between the speaker groups were about the same as the differences within each group. So linguistic categories don't map directly onto the conceptual groupings that underlie similarity judgments, suggesting that if conceptual groupings have essences, and if linguistic groupings have essences, they are not the same. Sloman and Ahn (1999) provide another example of a dissociation between linguistic and non-linguistic task performance.

In sum, the cross-linguistic data indicate that we can't appeal to labels to decide what comprises the category whose members are believed to share an essence (see Rips, 2001, for a contrasting view). This is a problem for essentialism because there's no clear alternative to appeal to. Obviously, we can't appeal to objects' functions because that would quickly become circular for the intended function view (the essence of the group of things intended to pound nails is that they were created to pound nails). Can we appeal to objects' inductive potential (as, e.g., Mak & Vera, 1999, and Mandler & McDonough, 1998, do)? Are there conceptual groupings consisting of all and only those things that support similar inductions? The problem here is the lack of an independent basis for determining the relevant inductions. We can't use just any induction because patterns of induction are property-specific (Heit & Rubinstein, 1994; Sloman, 1994). A grouping of carnivores vs. insectivores supports some inductions (if an owl eats it, that makes it more

likely that a bobcat eats it), but not other kinds of inductions (just because an owl has sesamoid bones doesn't mean a bobcat does). Other groupings (e.g., birds vs. mammals) might support the latter but not the former. So patterns of induction don't provide a unique segregation of objects into kinds.

One might object to our argument as applied to the domain of our example, natural kinds, because natural kinds exhibit clusters of correlated properties (Rosch & Mervis, 1975). To the extent that these correlated properties result in the perception of well-articulated clusters of instances separated by large gaps, then the resulting clusters will support inductions better than any classification that cross-cuts the clusters. What the example shows is that inductive power is gained by considering multiple classification schemes for the same set of objects, and human induction shows this more flexible character. In any case, artifacts show less clustering of ascribed properties than natural kinds do (Keil, 1995; Malt et al., 1999).

Can we determine the relevant groupings by appealing to causal structure (an object belongs to a particular conceptual group if it obeys certain causal laws) as Rips (2001) suggests? Perhaps, but not in a way that is independent of the presumed essence. Essences are defined as fundamental causes, so essentialists already appeal to causal structure to define the notion of essence. Therefore, they cannot also use causal structure to determine what the relevant grouping is because that, again, would be circular. The theory would be predicting merely that the beliefs people hold about the causal structure of objects determine the sets of things that they believe share a causal structure.

In conclusion, the view that intended function constitutes the essence of artifacts does not seem to help explain how people determine an artifact's kind. We next consider a different definition of essence to see if that withstands scrutiny.

Essentialist views of artifacts: II. Creator's intention as essence

Bloom (1996) has offered an alternative essentialist view of judgments of category membership: that people take an object's category to be whatever the maker intended it to be. They classify something as a chair if they believe it was created to be a chair. On this view, people take the maker's intended category membership as the essence of the category. This theory is not vulnerable to Problems 1, 2, or 3 above because it doesn't rely on function to define the essence. The fact that functions or intended functions aren't regularly given priority in naming decisions isn't a problem for the same reason. This view does face some difficulties though.

1. Without some independent notion of category identity, the view can't be tested.

Like the first version of essentialism, testing this view is faced with the problem of circularity. To determine if belief in an essence is the determinant of judged kind membership requires that the essence and category be independently defined. Beliefs about intended function can't serve to define the relevant categories for the reasons stated in Part I. Nor can we appeal to inductive potential or causal structure, also for reasons given above.

Bloom (1996) often appeals to names, implying that a category is the set of things that a label applies to. But this leads to another problem discussed above: Cross-linguistic differences imply that the essence can't be associated with a linguistic label unless the essence is understood as something that is language relative. Could beliefs

about the maker's intended linguistic category membership for different objects be different in different linguistic communities? Presumably, an American manufacturer of two objects could intend for them both to be labelled "bottle" and a Spanish maker of the same two sorts of objects could intend for them to be called by two different names, e.g., "tarro" and "mamadera." In this case, Bloom would have to say that speakers of the different languages (somehow) understand these different intentions for the same objects. But how would speakers figure out the different intentions? It can't be via the name they would use for the objects, because that is the thing to be explained. Furthermore, in Malt et al. (1999), the very same set of objects were named differently by speakers of different languages. So it couldn't possibly be the case that linguistic differences can be reduced to (correct) understanding of different creators' intentions: The events of creation were identical, yet the linguistic categories assigned were different. At best, the linguistic differences might be attributed to mistaken inferences of different intended membership by speakers of different languages. But then the basis for their differing inferences is the crucial point to account for, and it remains unexplained.

Finally, the cross-cultural dissociations between language and similarity reported by Malt et al. (1999) must somehow be accommodated by this view. One possible resolution is that people across cultures share beliefs about makers' intended membership and these beliefs serve as the essence, not of the linguistic categories that vary across cultures, but of the conceptual groupings that our similarity judgments suggest are close to universal. Bloom (1996) is blocked from making this move, however, because his goal is to explain what objects are called, not their non-linguistic groupings. In any case, nonlinguistic groupings themselves can vary with the nonlinguistic task used (see below).

2. When creator's intended category membership is pitted against other aspects of the objects such as physical structure, kind judgments are sometimes governed by the other aspects.

A second problem for this theory of artifact categorization was revealed by Chaigneau (2002; see Barsalou, Sloman, & Chaigneau, in press). He considered cases where people have privileged access to the creator's intended category membership by being told it. He pitted such knowledge against other aspects of the object and considered their relative importance in judgments of the appropriateness of a label (as well as judgments of the object's function and the object's causal efficacy). He described scenarios in which one of four aspects of a common object (e.g., a mop) was unusual (different from that of a normal object): its intended category membership, its physical features, the agent's intended use for the object, or the actions performed with the object. To create an unusual intended category membership, Chaigneau described an accidental creation scenario to participants in which the object was not created for any particular purpose. For example, Jane accidentally attached a bundle of thick cloth to a 4-foot long stick and John subsequently used it to wipe up a water spill. To create unusual physical features, a scenario was described in which the object was created to be a mop but was not normal physically. For example, John wiped up water with an object that was made to wipe up spilled water, but the object was a bundle of plastic bags attached to a 4 foot long stick. In the two remaining mop scenarios, the object was intended to be a mop. In one, the agent used the artifact normally but unintentionally (John accidentally pressed the object against a water spill). In the other, the agent performed unusual actions with the object (John pressed the wooden stick rather than the cloth against the water spill).

What Chaigneau (2002) found was that the creator's intended category membership was not the most important variable for any rating task. Most relevant here, changing the creator's intended category membership had less effect on judgments of the goodness of a name than changing the object's physical structure did. A bundle of plastic bags attached to a stick was judged to be less appropriately called a "mop", even though it was intended to be a mop, than an object created by accident that served the wiping up water function very well. People's choice of names were clearly not primarily guided by intended category membership.

These results seem to contrast with those of Gelman and Bloom (2000), Keil (1989), Matan and Carey (2001), and Rips (1989). Rips, for example, asked people whether an object created to be a lampshade but with the physical structure of an umbrella was more likely to be an umbrella or a lampshade. Most people thought it was more likely to be a lampshade, a result that would seem to favor the importance of historical creation over physical structure for kind membership. Those studies differ from Chaigneau's (2002) in several ways. For one, Chaigneau specified the physical structure of his objects precisely (e.g., a 4 foot stick with a bundle of plastic bags attached to one end). In contrast, Rips's descriptions left room for interpretation. For example, one description said the object was "a collapsible fabric dome. It consists of multicolored waterproof fabric stretched taut across six metal struts radiating from a central post in the dome. The metal struts are jointed so that they may be folded and this allows the fabric dome to be collapsed. When fully extended the dome is about three feet wide. [The creator] intended for this object to be used with the inside of the dome facing up as an attachment to ceiling light fixtures..." Although the description includes a lot of detail and the object is clearly similar to an

umbrella, it does not say explicitly that the object has the physical structure necessary to function as an umbrella (e.g., is the fabric permeable?). In contrast, it must have the physical structure necessary to function as a lampshade, because that's what it was intended to be. In Rips's control condition, in which the object clearly does have the physical structure necessary to serve as an umbrella, people were more likely to consider it an umbrella than a lampshade. Hence, the experimental description had just enough ambiguity to allow physical structure to be interpreted in a way consistent with the rest of the story, with the result that the experiment, unlike Chaigneau's, did not directly pit a fixed physical structure against creator's intention. In general, participants' judgments seem to be more closely aligned with the objects' physical structure, or inferred physical structure, than with any other aspect of the object, including its intended category membership.

Like those of Malt et al. (1999) and Sloman and Ahn (1999), Chaigneau's (2002) data show a divergence between naming and conceptual judgments (also see Gennari, Sloman, Malt, & Fitch, 2002). The various scenario aspects showed a different pattern of effects on judgments of naming and judgments of function (as well as judgments of causal efficacy). For example, the agent's actions had more influence on naming than function ratings. This provides further evidence against the possibility that conceptual groupings could be grounded in linguistic ones.

Conclusions regarding essentialism

All in all, psychological essentialism suffers from several critical problems when applied to artifacts. One virtue of essentialism, however, is that it seems to offer a way to think about how people are able to have *modal* beliefs; that is, beliefs about what would

be true about an object even if it had properties other than those it actually does have (Rips, 2001). Consider an animal that has all the properties of a giraffe except that it has stripes instead of spots. If you believe that such a beast would be a giraffe, this is a modal belief because presumably no such animal actually exists. What supports this belief? Not direct experience because one can't have experience of things that don't exist. The essentialist answer is that our belief in essences is what makes such beliefs possible. The belief that giraffes have some true, underlying nature that imparts kind identity and that this nature is the cause of giraffes' observable properties would lead one to conclude, given enough causal knowledge, that spots and stripes are merely observable effects of more fundamental, essential properties. As long as our hypothetical animal retains the essence of the category, it should be judged a category member regardless of its appearance.

But Rips (2001) points out that essentialism is not the only theory available to explain the existence of modal beliefs. A different type of theory appeals, not to intrinsic properties like essences, but to extrinsic relations that objects have with their environments. Modal beliefs could be supported by knowledge of the role that objects play in causal interactions with other things. Something with stripes would still be a giraffe because having stripes would not change the causal relations between the animal and its niche. It could still be its parents' progeny, it could still breathe oxygen, it could still eat leaves high off the ground, etc. And these beliefs about objects' causal interactions with their environments don't depend on beliefs in essences; in fact, often they're mediated by very superficial properties.

The minimalist view

Strevens (2001a) makes an argument against essentialism related to Rips' (2001). He points out that essentialism explains categorization and inductive judgments of biological and chemical categories by positing that people (a) have a belief that the relevant category has an essential property and (b) have causal knowledge about the category. Strevens' argument is that only (b) is actually relevant to explaining what people do on categorization and induction tasks.

To understand the flavor of Strevens' argument, first consider his analysis of how essentialism explains an experiment reported in Keil (1989) (Strevens actually considers three variants of essentialism). Keil told both children and adults of an animal (e.g., a raccoon) that had been transformed cosmetically to appear just like a different animal (e.g., a skunk, by adding a distinctive odor, white stripe down its back, etc.). When asked if the animal is a skunk or a raccoon, 2nd graders and adults tended to call it a raccoon (though kindergartners tended to call it a skunk). The essentialist account of the older participants' responses posits that people use causal knowledge to decide which of an object's features are most likely to be direct causal consequences of being a category member. The most direct causal consequences have the most influence in categorization decisions because they are the least defeasible. Explaining how a raccoon could have a stripe on its back is easier (because someone could have painted it there) than explaining how a skunk could have begun life as a raccoon. Because the causal inference explaining how this strange creature could be a raccoon is easier to construct than the one explaining how it could be a skunk, we call it a raccoon. In general, properties are given more weight in the categorization decision to the extent they are causally central (Ahn, 1998)

because the ease of explaining away a property is inversely proportional to its centrality (Sloman, Love, & Ahn, 1998). Note that essence plays no role in the explanation of the data. Keil examined not only animal transformations but also transformations of one artifact to another (e.g., a kitchen pipe that is turned into a flute). He found that all participants were more likely to say that a kitchen pipe with holes that can be used to make music is more likely to be a flute than a kitchen pipe. This can also be explained by appealing to defeasibility conditions derived from a causal analysis of object properties. An explanation for why something with holes that can be used to make music would be a flute is easier to generate than for why it would be a kitchen pipe.

Essentialism provides a related account of how inductions about generic, unfamiliar properties are made. A property is projected from one entity to another to the degree that a causal analysis leads to certainty that the properties of the two entities have a common source (a claim like this is made, for example, by Gelman, 1988).

Strevens (2001a) makes two important points. First, he argues that causal knowledge comes in the form of universal categorical assertions that he calls K-laws. K-laws have the form "All Ks have P" in which K is a natural kind and P is an observable property (e.g., all raccoons have raccoon parents by virtue of the causal process of reproduction). Second, he argues that K-laws do all the work in explaining categorization and induction, that any further assumption that categories have essences is superfluous and unnecessary. He therefore posits a non-essentialist theory of categorization that he calls minimalism, that categorization and induction are driven by knowledge of K-laws (no essence required).

Strevens (2001b) is not willing to extend his argument to artifacts. His unwillingness could arise for several reasons. We consider four possibilities: i. There is a dearth of clear causal laws governing artifacts. Many properties of artifacts are arbitrary, unconnected to any causal system inherent to the object. The color and texture of telephones, refrigerators, cups, etc. are independent of the other properties of those objects. More generally, many properties of artifacts concern aesthetics and design and can be selected with minimal consideration of the object's causal properties and therefore are unrelated to a causal system specific to the object. ii. Relatedly, artifacts -- individually or in groups -- support fewer inductions than natural kinds because their properties do not occur in such tightly clustered sets. For instance, the fact that a screwdriver has some property or component is a weak reason to project the property to any other artifact because the property or component is likely to be specific to activities involving the screwdriver, or it may be there for aesthetic reasons. iii. Many artifacts have little or no internal structure. The structure of a plate or a table is simple and doesn't lend itself to much causal analysis. iv. Finally, members of natural kinds tend to evolve or develop in a specific niche, governed by fixed causal laws. In contrast, artifacts can emerge in multiple environments in multiple ways and can serve many roles, sometimes simultaneously. To illustrate, jars have emerged in many forms (ceramic, tin, glass) in many societies, often to serve different functions (to store wine, to carry water, even to bake cakes). With so many roles to play, there may not be a fixed set of causal laws governing them.

But all of these arguments concern matters of degree, not fundamental differences between artifacts and natural kinds. We consider each possibility in turn. i. Many

properties of artifacts are arbitrary, but many are not. The color of a telephone may be arbitrary, but it should have a microphone and a speaker, and these should be positioned to allow use by someone with a mouth and an ear in the specific locations that one finds them on the human head. Pens should be graspable, and they should extrude ink at a constant rate.

ii. Most artifacts support fewer inductions than natural kinds, but all support some. In fact, Farrar, Raney, and Boyer (1992) and Sloman (1998) found no difference between the number of inferences drawn from artifacts and natural kinds. Even if all one knows is that an object is a paperweight, one can induce that its mass is within a certain range. Moreover, some artifacts allow more inductions than some natural kinds. Knowing the properties of cars offers many hints about the properties of trucks. Not many of the properties of clouds, a natural kind, generalize to other entities.

iii. Most artifacts don't have much internal structure, but some have a great deal (cars, rocketships, computers, clocks, player pianos, etc.). More importantly, artifacts have critical causal structure, namely the relations between their parts and operations and the function they serve to the external agent who uses them (or even just appreciates them).

iv. Although some artifacts have a variety of functions, others do not (e.g., an atomic clock). And even for those that do, a multiplicity of functions doesn't imply that the object isn't governed by a fixed set of causal laws. The laws may vary with the function, but that just means they're context-specific, not absent. Often a minimal set does carry over from context to context. In the example above of a jar, the causal laws related to containment apply throughout. In sum, we see no principled reason to limit any theory of how people categorize to natural kinds or artifacts alone. The two domains differ in central tendency, but they overlap (Keil, 1989; Markman, 1989).

If minimalism does apply to artifacts as much as natural kinds, then Strevens' (2001a) argument against essentialism should apply to artifacts as much as to natural kinds. The claim of such an argument would be that, once assumptions about causal knowledge are made, no assumption about essence is necessary to explain how people group or make inductions about artifacts. Presumably, the relevant causal knowledge for artifacts would be twofold. First, it would concern how creators' intentions get realized in physical media such that objects are created to perform certain functions. Second, it would concern how agents use objects to actually perform particular functions. Once this causal knowledge and its relation to judgment is spelled out, no notion of essence -- of a true underlying nature that confers kind identity -- does further explanatory work because there's no more work to do.

How does the minimalist view fare with artifacts?

Minimalism describes the causal knowledge enabling categorization in terms of K-laws, universally quantified relations between kinds and properties. Applying the theory to artifacts, an example of a K-law might be "all hammers pound nails," where *hammers* refers to a kind and *pound nails* to a property that's a causal effect afforded by being a hammer.

But do we need to assume that the causal beliefs people use to make judgments and to reason about objects are beliefs about a relation of causal properties to kinds per se? A theory even more minimal than minimalism may be sufficient to explain most nonlinguistic judgments about artifacts. Such a superminimalist theory would dispense not only with the assumption that judgments about kinds are determined by an essence as

minimalism does (Strevens, 2001c), but also with the assumption that causal beliefs about artifacts are organized around kinds at all.

One suggestion of the Malt et al. (1999) work is that common containers don't have a single natural partitioning, but rather different tasks and different languages partition them in different ways. These partitionings depend on the demands of the specific categorization task and historically derived conventions about how to perform the task. Other studies have shown that inductions are not governed by a fixed category structure but rather depend on specific task demands (Heit & Rubinstein, 1994; Sloman, 1998).

Such task relativity suggests that objects of the type we've studied aren't perceived as belonging to any one kind of thing; their classification depends on the purpose of the task. Different tasks have different goals and people excel at learning to attend to properties and structural relations that satisfy their goals. The relativity of category structure is consistent with McCloskey and Glucksberg's (1978) finding that people vary considerably in their category judgments from day to day. In their data, 25% of instances received a different category assignment (measured as the name given) from the same participant a month later.

In science, causal powers are carried by properties, not by objects per se. Diamond cuts granite because of their relative hardnesses. If a piece of granite were sufficiently hard, diamond wouldn't cut it. Indeed, causal laws in science describe relations between properties, not kinds (e.g., $F = ma$ can be construed as a causal relation amongst three properties that hold regardless of which object is in motion). The fact that science organizes causal knowledge around properties, not kinds, suggests – at minimum – that that's a good way to understand things. People may take advantage of this organizational

principle, at least for artifacts. In other words, even if causal knowledge isn't organized around kinds, causal laws may still govern judgment about artifacts (if it's impermeable to water, it can store liquids; if it's sufficiently flat, it can be used to serve food). But these aren't K-laws because they don't relate kinds to properties; instead they relate properties to properties.

Organizing causal knowledge around properties rather than kinds supports flexibility and task relativity because the properties relevant to a task can usually be selected and attended to with relative ease. Massive task relativity doesn't make as much sense for living things as it does for artifacts. The notion of kind is critical to folk classification and induction of living things, which tend to be consistent across cultures (López, Atran, Coley, Medin, & Smith, 1997; Malt, 1995; note however that McCloskey & Glucksberg's, 1978, data show equal variability across time in category judgments between natural kinds and artifacts and Sloman, 1998, showed systematic neglect of category structure in inductions over living things). This is a consequence of biology; many important generalizations apply to living things at levels more abstract than those that apply to artifacts (e.g., for all living things: if its parents are of kind X, then it's an X; if it's of kind X, then it has a particular physical structure). The kinds relevant to these abstract inference rules are stable enough, and the inferences are important enough in our daily lives, that they are part of our inventory of everyday inference rules.

These abstract inference rules that take kinds as arguments are more prevalent and provide more inductive power in biology than in the artifact domain. Relatively few abstract inference rules apply to artifacts, especially very general rules. Even apparently strong ones like "if it's a vehicle, then it transports people or things" admit of many

exceptions (junkyards are full of them). Another difference between natural kinds and artifacts, as we have already noted, is that many natural kinds, especially those encompassing living things, cluster more tightly in similarity space with larger gaps between clusters than artifacts do (though cf. Malt, 1994, on water). As a result, natural kinds are likely to show less divergence across tasks and cultures than artifacts.

Conclusion

We conclude that psychological essentialism and minimalism are both underdetermined by the evidence in the artifact domain. Psychological essentialism suffers from inadequacies in the characterization of essence either as intended function or in terms of creators' intention. Also, the task relativity of judgment makes the theory untestable because no independent, acceptable method exists to determine whether people assign an object to a "kind." Both types of theory make unnecessary assumptions, namely that categorization is determined by a belief in essences (essentialism) and that people group artifacts into stable clusters constituting kinds (essentialism and minimalism). These assumptions are unnecessary because both essentialism and minimalism require further assumptions about causal knowledge relating properties to one another to explain how people make classification and induction judgments, and those further assumptions are sufficient to explain the judgments.

Both essentialism and minimalism go wrong in assuming that people represent artifact kinds with a stable set of beliefs separate and distinct from the tasks used to classify objects. We propose that there are no fixed artifact categories in the head. Artifact categories have no fixed boundaries, even fuzzy ones. Of course, objects cluster in particular contexts; they are more or less similar to one another in those contexts.

However, to say that objects cluster in some conceptual space is not to say that objects must be understood as of one kind or another. The fact that judgments of similarity are notoriously labile means that an appeal to similarity can't impart confidence in the stability of category structure. Similarity judgments depend on a host of contextual factors (knowledge of the judge, the set being judged, the nature of the similarity judgment task; Sloman & Rips, 1998; see Medin, Goldstone, & Gentner, 1993, for a review). Induction tasks impose their own structure on objects. In the case of induction, that structure depends on a causal analysis of the specific property being projected. Other tasks may result in yet other groupings. For instance, it is an open question how the space of objects would be cut up by recognition memory. Confusions in recognition can be construed as a measure of similarity, but they can differ systematically from explicit similarity judgments (e.g., Gennari et al., 2002). One reason is that recognition is known to be sensitive to frequency, whereas similarity is less so.

Therefore, artifact categories depend on how the categories are elicited -- on the categorization task at hand. Naming is one categorization task that cuts up the space of artifact objects, but different languages do it differently. And the way each does it is conventionalized, depending on the specific history and structure of the language. Naming is also governed by the specialized purpose of language -- to communicate -- and the communicative context may have specific effects on the names people choose (e.g., Brennan & Clark, 1996; Malt & Sloman, 2001). Obviously people use language to learn about the properties of artifacts and indeed linguistic labels are sometimes given priority in the inductive process even by young children (e.g., Gelman & Markman, 1988). Presumably, this is because linguistic cues are extremely effective pointers to the

existence of shared structure between objects. But this doesn't mean that linguistic categories are mirrored by nonlinguistic mental representations that underlie object knowledge. It merely means that words used appropriately in context can be effective pointers to nonlinguistic structure. Surely nonlinguistic structure exists. For example, artifacts share more or less causal structure with one another. And such structure is indeed correlated with the names we give things. But the correlation is far from perfect because we do different things with it as required by specific tasks. Moreover, tasks add their own constraints. We have argued that causal knowledge is organized around property relations and not around kinds for exactly that reason: so that different tasks can pick out the relevant bits of knowledge. A well-designed conceptual system shouldn't have fixed boundaries when the knowledge plays a variety of different roles. Allowing different tasks to partition objects differently according to their demands enhances the system's flexibility.

Our conclusion is inconsistent with theories other than just essentialism and minimalism. It is inconsistent with any theory that assumes that kinds are fundamental, such as theories that assume defining features for category membership or theories that explicitly impose category boundaries in their representation (e.g., Ashby, 1992). It is also inconsistent with exemplar theories that assume that exemplars are stored with a single category label because the labels impose an implicit boundary (Kruschke, 1992; Nosofsky, 1988). A different type of exemplar model would remain tenable though, one that does not store a label with each exemplar. Labels could either be stored independently or multiple labels might be stored with each object. Such a representation could allow category boundaries to vary with the task by differentially weighting

exemplars to generate a response, with the weighting depending on task demands. Of course, how causal structure can be abstracted from an exemplar representation remains an open (and difficult) question.

Our claim that artifacts don't come in kinds violates a strong intuition. It seems right to say that a hammer is a hammer, it's not a nail; the two objects are of fundamentally different kinds. How can we say otherwise? Relatedly (but not equivalently), how can we say that essences don't matter when people feel so sure that they do? Remember that the cross-linguistic data we have described indicate that the linguistic intuition must be separated from the conceptual one. We agree that an object that all English speakers would call a "hammer" should almost never be called a "nail." Not only would that violate Gricean maxims of cooperativeness, nobody would know what you're referring to. Our claim isn't that objects don't have better or worse names in a particular language. Patterns of naming do yield linguistic category boundaries. However, those linguistic categories are a function of particular linguistic and cultural histories (Malt et al., 1999) and objects' roles in systems of relations (Barsalou et al., in press; Markman & Stillwell, 2001; Rips, 2001) as much as of inherent properties of the objects. Our claim is that talking about an artifact's kind does not help us much to explain how people perform everyday conceptual classification and inductive judgment tasks. This claim can be true even if people believe that artifacts come in kinds. People can hold beliefs about essences and beliefs about how essences relate to kind membership without those beliefs having any causal relation to judgments that put object knowledge to use.

Why do people have such strong intuitions about kinds? We suspect that the intuition derives at least in part from the ease and automaticity with which people represent

knowledge using language. People do have knowledge about artifacts, such as causal knowledge about how properties relate to other properties, and giving an object a name is often intended to convey that knowledge. If my uncle points at a machine across the street and says "that's a trencher," he is asserting that he knows it's for digging trenches, that he knows enough about its parts and their relations to determine its function, and that he knows the English convention for labeling the machine. It's a small step from the ability to use language to demonstrate our knowledge of objects to the belief that the successful use of a name for an object reflects a category membership that the object must hold by virtue of its properties.

References

- Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition, 69, 135-178.
- Ahn, W-k., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D. & Shafto, P. (2001). Why essences are essential in the psychology of concepts, Cognition, 82, 59-69.
- Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.). Multidimensional models of perception and cognition. Hillsdale, N.J.: Erlbaum.
- Barsalou, L. W., Sloman, S.A., & Chaigneau, S. E. (in press). The HIPE theory of function. To appear in L. Carlson & E. van der Zee (Eds.), Representing functional features for language and space: Insights from perception, categorization and development. New York: Oxford University.
- Bloom, P. (1996). Intention, history, and artifact concepts. Cognition, 60, 1-29.
- Brennan, S. E. and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1482-1493.
- Chaigneau, S.E. (2002). Studies in the conceptual structure of object function. Doctoral dissertation, Department of Psychology, Emory University, Atlanta, GA.
- Farrar, M. J., Raney, G. E., & Boyer, M. E. (1992). Knowledge, concepts, and inferences in childhood. Child Development, 63, 673-691.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. Cognitive Psychology, 20, 65 - 95.

- Gelman, S. A. & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. Cognition, *76*, 91-103
- Gelman, S. A. & Hirschfeld, L. A. (1999). How biological is essentialism? In D. L. Medin & S. Atran (Eds.), Folkbiology. Cambridge: MIT Press.
- Gennari, S. P., Sloman, S. A. Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. Cognition, *83*, 49-79.
- Hampton, J. A. (1995). Testing the prototype theory of concepts. Journal of Memory and Language, *34*, 686-708.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. Journal of Experimental Psychology: Learning, Memory, and Cognition, *20*, 411-422.
- Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.
- Keil, F. C. (1995). The growth of causal understanding of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Causal cognition: A multidisciplinary approach (pp. 234-262). New York: Oxford University Press.
- Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), Semantics of natural language. (pp. 253 - 355). Dordrecht, Netherlands: D. Reidel Publishing.
- Kronenfeld, D.B., Armstrong, J.D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J.W.D. Dougherty (Ed.). Directions in cognitive anthropology (pp. 91-113). Urbana: University of Illinois Press.

- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, *99*, 22-44.
- Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. Journal of Memory and Language, *38*, 1-27.
- López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. Cognitive Psychology, *32*, 251-295.
- Mak, B.S.K. & Vera, A. (1999). Motion-based categorization in young children. Cognition, *71*, B11-B21.
- Malt, B.C. (1994). Water is not H₂O. Cognitive Psychology, *27*, 41-70.
- Malt, B. C. (1995). Category coherence in cross-cultural perspective. Cognitive Psychology, *29*, 85-148.
- Malt, B. C. & Johnson, E. C. (1992). Do artifact concepts have cores? Journal of Memory and Language, *31*, 195-217.
- Malt, B.C. & Sloman, S.A. (2001). Beyond conceptual pacts: Enduring influences on lexical choice in conversation. Manuscript under review.
- Malt, B.C., Sloman, S.A., Gennari, S., Shi, M., and Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. Journal of Memory and Language, *40*, 230-262.
- Mandler, J. M. & McDonough, L. (1998). Studies in inductive inference in infancy. Cognitive Psychology, *37*, 60-96.
- Markman, A. B. & Stillwell, C. H. (2001). Role-governed categories. Journal of Experimental & Theoretical Artificial Intelligence, *13*, 329-358.

- Markman, E. M. (1989). Categorization and naming in children. Cambridge: MIT Press.
- Matan, A. & Carey, S. (2001). Developmental changes within the core of artifact concepts. Cognition, *78*, 1-26.
- McCloskey, M. & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? Memory and Cognition, *6*, 462-472.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. Psychological Review, *100*, 254-278.
- Medin, D. L. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), Similarity and analogical reasoning. New York: Cambridge University Press.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representation. Journal of Experimental Psychology: Learning, Memory, and Cognition, *14*, 54-65.
- Putnam, H. (1975). The meaning of 'meaning.' In H. Putnam, Mind, language, and reality: Philosophical papers, Vol. 2. Cambridge, England: Cambridge University Press.
- Rips, L. (2001). Necessity and natural categories. Psychological Bulletin, *127*, 827-852.
- Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, *7*, 573-605.
- Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. Cognition, *52*, 1-21.
- Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. Cognitive Psychology, *35*, 1-33.

- Sloman, S. A., & Ahn, W. (1999). Feature centrality: naming versus imagining. Memory & Cognition, *27*, 526-537.
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. Cognitive Science, *22*, 189-228.
- Sloman, S. A. Malt, B. C. & Fridman, A. (2001). Categorization versus similarity: the case of container names. In M.Ramscar, U.Hahn, E.Cambouropolos, & H.Pain (Eds.). Similarity and categorization. Cambridge: Cambridge University Press.
- Sloman, S. A. & Rips, L. J. (1998). Similarity as an explanatory construct. Cognition, *65*, 87-101.
- Stevens, M. (2001a). The essentialist aspect of naive theories. Cognition, *74*, 149-175.
- Stevens, M. (2001b). Only causation matters: reply to Ahn et al. Cognition, *82*, 71-76.
- Stevens, M. (2001c). Further comments on Ahn et al. <http://www-csli.stanford.edu/~stevens>.

Acknowledgements

We thank Sergio Chaigneau, Larry Barsalou, Art Markman, James Hampton, an anonymous reviewer and especially Daniel Weiskopf for insightful and helpful comments on the manuscript. This work was funded by NASA grant NCC2-1217 and NIMH Grant MH51271.