ELSEVIER

Cognition 60 (1996) 1–29

COGNITION

Intention, history, and artifact concepts

Paul Bloom*

Department of Psychology, University of Arizona, Tucson, AZ 85721, USA

Received 18 November 1994, final version accepted 6 October 1995

Abstract

What determines our intuitions as to which objects are members of specific artifact kinds? Prior research suggests that factors such as physical appearance, current use, and intended function are not at the core of concepts such as *chair*, *clock* and *pawn*. The theory presented here, based on Levinson's (1993) intentional-historical theory of our concept of *art*, is that we determine that something is a member of a given artifact kind by inferring that it was successfully created with the intention to belong to that kind. This theory can explain why some properties (such as shape) are more important than others (such as color) when we determine kind membership and can account for why certain objects are judged to be members of artifact kinds even though they are highly dissimilar from other members of the kinds. It can also provide a framework for explaining the conditions under which broken objects cease to be members of their kinds and new artifacts can come into existence. This account of our understanding of artifact concepts is argued to be consistent with more general "essentialist" theories of our understanding of concepts corresponding to proper names and natural kind terms.

1. Introduction

What underlies our intuitions as to which things in the world are chairs, clocks, and pawns? To put it differently, adopting the terminology of cognitive psychology, what is the nature of our concepts *chair*, *clock*, and *pawn*? This paper concerns the nature of concepts that correspond to kinds that exist through human agency ("artifact kinds"); these are distinct from natural kind concepts such as *water* and *lemon*.

It is unlikely that our intuitions about kind membership are based solely

* Fax: 520 621 9306; e-mail: bloom@u.arizona.edu

SSDI 0010-0277(95)00699-0

on physical features. We do know what a typical chair looks like; sighted people can usually recognize chairs through vision, and blind people can do so through touch. But perceptible qualities such as being of a certain size and shape cannot entirely constitute our concept of *chair* since this concept includes entities of radically dissimilar appearance, such as beanbag chairs, basket chairs, deck chairs, chairs for dolls, chairs shaped like hands, chairs suspended from ceilings on chains, and so on.

This diversity of appearance is typical of artifact kinds. Clocks do not share any common shape, size, or texture – there exist analog clocks, digital clocks, and talking clocks shaped like ducks that are activated with the press of a button. Pawns, since their function is quite unrelated to their material nature, can have virtually any physical property so long as they are perceptible, moveable, and can fit on a chess board. Most typically, they are small white and black objects, but they can also be green marble slabs, pennies (used in a pinch if pieces from a set are lost), and people dressed in medieval costumes (as in some outdoor chess games).

Although there exist cases in which it is unclear whether or not an object belongs to an artifact kind (e.g., Lakoff, 1987). these are not necessarily the same cases in which the object has an atypical appearance. A beanbag chair, a talking duck-shaped clock, and a giant green pawn are certainly unusual examples of their kinds, but this does not cause us to question their status as chairs, clocks, or pawns. This suggests that there exists a dissociation between how typical something is as a member of a kind and the extent to which it *is* a member of a kind. This dissociation is found in other domains as well. For instance, 3 is consistently judged by subjects to be a better example of an odd number than 447, even though the same subjects will insist that *odd number* is an all-or-nothing category, and so it makes no sense to judge some numbers as odder than others (Armstrong, Gleitman, & Gleitman, 1983).

The use to which an object is put is often relevant to the artifact kind it belongs to. Chairs are typically what we sit on, clocks are what we tell time with, and pawns are pieces that move in a certain circumscribed fashion on a chess board. Nevertheless, most people would judge that a broken chair that nobody could sit on was still a chair, and a clock that could no longer tell time was still a clock. Similarly, sitting on a desk does not turn the desk into a chair, and mistakenly moving a pawn diagonally across the board does not turn it into a bishop. Part of our understanding of artifacts, then, is that something can belong to a kind even if it is not used in the manner associated with typical members of that kind, and something can be used in the manner associated with typical members of a kind without being understood as belonging to that kind.

The above considerations suggest a more plausible theory of artifact concepts, which is that they correspond to kinds of entities that are intended to have the same use – a common intended function. Although not all chairs can be sat upon and desks can be sat upon, chairs and not desks are typically created with the intention that people sit on them. Perhaps this is why a broken chair is still a chair and a desk that is sat upon is not a chair. Under this view, what an object looks like and what is done with it are relevant only insofar as appearance and current use are reliable cues to its intended function. For instance, Rips (1989) found that if an object that is umbrella-like in appearance is described as being designed to serve as a lampshade people view it as a lampshade, not an umbrella (see also Hall, 1995), and Keil (1989) obtained similar findings of function overriding appearance in the domain of tools. This has suggested to many scholars that the psychological "core" of artifact concepts is that their members share a common intended function (see Malt and Johnson, 1992 for a review).

Such a proposal runs into serious problems, however, and these are reviewed in the section below. The remainder of this article will present a theory of artifact concepts that avoids these problems, one based on a previous proposal that has been made in the domain of art. A precursor to this theory will be introduced first for a subclass of artifacts – representational pictures – and then the theory will be presented for artifact kinds in general. This article concludes by relating this account of artifact concepts to more general theories about the nature of concepts, including definitional theories, prototype theories, and essentialist theories.

2. Problems with function-based accounts of artifact concepts

The main problem with the proposal that function underlies our understanding of artifact kinds is that, as function is normally defined, having the right function is neither sufficient nor necessary for something to be a member of an artifact kind.

The most systematic evidence in this domain emerges from an intriguing set of experiments by Malt and Johnson (1992). They started by collecting subjects' intuitions about the physical and functional features of several common artifact kinds. For instance, most subjects thought that the physical description "wedge-shaped, with a sail, an anchor, and wooden sides" and the functional description "manufactured and sold to carry one or more people over a body of water for the purposes of work or recreation" both applied to the kind *boat*.

Malt and Johnson then gave subjects descriptions composed of the functional features that correspond to specific artifacts, but with atypical physical features for these artifacts. For instance, they would be told about a thing which is "manufactured and sold to carry one or more people over a body of water for the purposes of work or recreation" (functional features associated with boats) but which is "spherical and made of rubber, is hitched to a team of dolphins, and has a large suction cup that can keep it in one place" (physical features not associated with boats) and they would be asked "Is this thing a BOAT?" In these instances, about two-thirds of the subjects denied that the object was a member of the artifact kind – while virtually everyone accepted the object as belonging to the artifact kind when it was described with the appropriate functional features *and* the appropriate physical features. This finding suggests that functional features are not sufficient to determine artifact kind membership in the judgements of naive subjects.

In another experiment, subjects were given physical descriptions appropriate to the kind (e.g., "wedge-shaped, with a sail, an anchor, and wooden sides") but inappropriate functional descriptions, sometimes related to the function associated with the artifact kind (e.g., "manufactured and sold as a holding area for dangerous criminals or persons in exile by detaining them a certain distance off-shore") and sometimes inconsistent with it (e.g., "manufactured and sold for collecting samples of marine flora and fauna under sterile conditions, and is totally mechanized so that no people are allowed onboard under any circumstances"). In such cases, over half of the subjects said that the thing *did* belong to the artifact kind, suggesting that functional features are also not necessary for artifact kind membership.[1]

Landau (1994) reports a similar finding: children who are taught names for novel artifacts will extend the names to further objects of the same shape, but not to further objects that share the same function. Such results support the claim that functional features are not at the core of artifact kinds, and Malt and Johnson (1992) tentatively favor a family resemblance theory in which different factors, both physical and functional, conspire to determine our intuitions about kind membership.

There are a couple of objections that one could raise concerning this interpretation of the experimental results. One concerns the finding that when people were told about objects that looked like boats, but lacked the typical functions of boats, they usually said "yes" to the question "Is this a BOAT?" The problem here is that subjects could give this response even if they do not believe that the objects actually do fall into the class of boats, since there is a difference between which things people use the word "boat" to describe and which things they actually view as boats. For instance, subjects would also agree that a drawing of a boat "is a boat", but this surely cannot entail that they believe that boats can be two-dimensional patterns of ink on paper. It instead reflects the fact that a word that refers to X can also be used to refer to a visual representation of X (Jackendoff, 1992).

This renders the nature of the "yes" responses unclear. If people were

[1] As an anonymous reviewer points out, however, some of the "physical descriptions" used by Malt and Johnson were themselves function-laden. Words like "sail" and "anchor" are not neutral physical descriptions of parts. Instead they have associated functions, and are also understood as being associated with boats. These factors might make subjects more likely to view objects described in this manner as actually being boats, and thus might exaggerate the extent to which categorization is determined by non-functional physical considerations.

agreeing that the predicate "is a boat" applied to the objects described as not being able to carry passengers because they believed these objects to fall into the class of boats then this supports Malt and Johnson's conclusion that function is not necessary for subjects' judgments about artifact kind membership; physical features can be sufficient. But if the subjects were saying "yes" because they were agreeing that the object *looks like* a boat or was *intended to depict* a boat, then this conclusion does not follow. Bloom (in press) and Soja, Carey, and Spelke (1992) discuss this concern more generally with regard to the appropriate interpretation of experiments that purport to show that children show a strong "shape bias" when learning new words (e.g., Landau, 1994; Landau, Smith, & Jones, 1988).

A different concern is that the functions used in the Malt and Johnson study, although they were clearly associated with the artifact kinds, might not be the critical ones. It is conceivable that if the correct functions were used, there would be no dissociation between possession of the function and membership into the artifact kind: All and only objects with the correct function for boats would be categorized as boats.

But what could these correct functions be? One might suggest, for instance, that the actual function of boats is less specific, such as "to carry or suspend objects over or above water." If so, then the purportedly "non-boat" functions provided by Malt and Johnson *are* consistent with the objects being boats, and it is compatible with a function-based theory that people are willing to say that objects with these functions are boats. As Malt and Johnson point out, however, positing this broad sort of function fails to distinguish artifact kinds at the appropriate level of precision. They note that the function "to carry or suspend objects over or above water" also includes rafts, life preservers, and cruise ships, and thus this function cannot serve as the core of a concept *boat* that is distinct from these other concepts. Similarly, it is true that chairs are usually designed for people to sit upon – but benches, stools, and sofas are also designed for this purpose. Consider also cups and bowls; it seems implausible that there is a function X and a different function Y such that we view cups as all and only those objects that have function X and bowls as all and only those objects that have function Y. What could such functions be? More likely, as Labov (1973) has argued, we make the cup/bowl distinction primarily on the basis of *shape*, not function.

As a modified proposal, one might suggest that function defines *classes* of artifact concepts – such as the class including rafts, life preservers, cruise ships, and boats, all of which were created to carry objects over water – while other factors, such as physical appearance, distinguish between these different kinds. But even this might be too strong. A more general problem with the functional view is that people can construct artifacts without intending them to be used at all. There is nothing incoherent about someone creating a boat without any desire that it end up in water, or even creating a boat with an express desire that it never end up in water. We would still view it as a boat, a fact that seriously undermines any function-based theory of artifact concepts.

3. Representational pictures and the concept of art

One type of artifact that has received considerable attention is that of intentionally created two-dimensional visual representations that depict entities in the world, such as sketches, cartoons, caricatures, paintings, and drawings. For the purposes here, these can be collapsed into the single category of *pictures*.

What is our notion of the artifact kind *picture of a dog* that distinguishes it from related kinds such as *picture of a cat*? Just as with chairs, clocks, and pawns, there is no physical property that all and only pictures of dogs possess. Some are tiny black-and-white sketches; others are huge abstract drawings. Some look like dogs (in the sense that one might even mistake them for actual dogs in dim light); others are diffuse smears of color that do not readily call dogs to mind.

One proposal is that our intuitive understanding of this kind is that all and only pictures of dogs were created with a specific kind of intention. This is plausible, but only if we choose the right intention. Proposals such as "the intention to represent an animal with courage and loyalty" or "the intention to represent something with four legs and a tail" cannot work, since a picture of a dog can be created without these particular intentions (one can intend to represent a three-legged scoundrel of a dog and it would still be a picture of a dog) and a picture can be created with these intentions without being a picture of a dog (it could be a picture of a courageous and loyal cat).

More plausibly, our naive notion is that a picture of a dog is a picture created with the intention to represent a dog. This distinguishes it from a picture of a cat which is the result of an entirely different intention – to represent a cat. In general, we construe the extension of the kind *picture of an X* as all and only those pictures created with the intention to represent an X.

This is not a proposal about the metaphysics of visual representation, it is instead a claim about people's understanding of pictures, our untutored intuitions as to how they relate to creator's intention. If we see someone draw a picture of a person and we do not know who it is a picture of, we are likely to ask the artist – we assume that he or she should know. Similarly, we accept that 3-year-olds draw pictures of their dogs, mothers, and houses, and the status of these pictures is rarely because they resemble dogs, mothers, and houses (see also Millikan, 1993). We also accept the possibility of bad representations; my picture of a dog might look more like a cat, but this makes it a bad picture of a dog, not a good picture of a cat. Similarly, if one wishes to imitate Bogart but sounds more like Mr. Smith across the block, it is a bad imitation of Bogart, not a good imitation of Mr. Smith.

We do not have psychic access to the intentions of others, however. This raises the question of how we can identify something as a picture of a dog just by looking at it, or as an impression of Bogart just by hearing it. We do so through a set of assumptions – plausibly viewed as a "naive theory" in the sense of Carey (1986) – as to how a person's beliefs and desires are expressed through her behavior and, in particular, through her acts of creation.

How do these intuitions apply in the domain of representational pictures? When we look at a picture and categorize it as belonging to the kind *picture of a dog*, what leads us to conclude that it was intended to represent a dog? There are indefinitely many ways that this could take place. The artist could tell us her intention in a sincere fashion. The picture could be titled appropriately, e.g., "My dog" (see Levinson, 1985 on titles). We could see the artist staring at a dog while creating the picture. But one particularly important cue is when we infer that the picture was intended to represent a dog – and is therefore a picture of a dog – by virtue of what it looks like.

In particular, it is a reasonable inference that something is a picture of a dog if it looks like a dog.[2] The logic behind this inference can be spelled out as follows: (i) A picture that looks like X is likely to have been created with the intention that it look like X; (ii) The usual reason why someone intends to make a picture that looks like X is because this is a good way to create something that will be recognized by others as representing X; and (iii) the intention to create something that will be recognized by others as representing X is normally associated with the intention to represent X. Thus, a picture that looks like X is normally the result of the intention to represent X, and is normally a picture of X.

All of the above qualifications are crucial, as there are many cases where a representation does not look like X but nevertheless represents X. Modern art is the most philosophically examined (e.g., Danto, 1981, 1992; Davies, 1991). Some artists will intend to create something that represents X but do not wish others to infer, or at least not easily, that it is a representation of X. In fact, they might choose to purposefully flout the

[2] Our judgments about which pictures look like which objects are not based entirely on "bottom-up" perceptual factors; background knowledge about conventions of art or the nature of representation in general also plays some role (Goodman, 1976). Nevertheless, some pictures actually share perceivable properties with their referents, and this is the sense in which pictures can be said to "look like" what they refer to (see Hagen, 1986). Note also that this capacity to identify some pictures on the basis of what they look like is unlearned (Hochberg & Brooks, 1962), indicating that an understanding of arbitrary conventions of art cannot be an essential aspect of picture recognition.

In fact, as Ittelson (in press) points out, many current theories of perception that purport to be about real-world object recognition (e.g., Biederman, 1987) are developed and tested on their capacity to recognize two-dimensional representations. Implicit in these theories, then, is the premise that our perception of pictures of objects of kind X and our perception of objects of kind X shape important properties, and a proper characterization of this overlap might explain why some pictures of X are thought of as resembling, or looking like, X.

conventions associated with the above generalization, so as to make a statement about art itself. Other artists do wish to inform others that the picture represents *X* but choose other ways to do so; they might have a helpful title or include the name of *X* (e.g., "DOG") within the picture itself (political satirists often use this technique, and it is useful for cases in which viewers cannot be assumed to know what the represented objects actually look like).

Also, as noted above, there is bad art. Bad artists sometimes wish others to believe that something is a representation of *X* and wish to do so by making it look like *X* – but fail due to incompetence. Someone might look at my picture (a clumsy line-drawing which looks vaguely like a cat) and guess that I intended to draw a cat – but this would be wrong. I intended to draw a dog; I just did a poor job at it. In this example, the ineptness of the picture makes it easy to attribute a range of different intentions to the artist, as it gives the impression that he is the sort who suffers a lot of slippage between intention and result.

The proposal thus far is that visual representations are categorized through what we infer to be the intentions of their creators. We view a picture of *X* as the result of the intention to generate a visual representation of *X*, and we categorize pictures through the application of a set of assumptions about the relationship between the appearance of visual representations and the intentions of those who create them. This proposal only applies to pictures that result from intentional psychological processes (drawing, painting, etc.), and not to visual representations such as photographs. If a camera is pointed at Fred in good light and a rock hits it and the shutter snaps, the result is a photograph of Fred, regardless of the lack of intention. Although the aesthetic qualities of photographs can be governed by intentions that are every bit as rich as those that underlie paintings, we tend to view their representational properties as (at least in part) the result of a non-intentional causal mechanism – the photograph is *of* whatever the camera is pointed at. Drawings and similar kinds of pictures, in contrast, are created through purely intentional processes, and our intuitions about these processes dictate how these are categorized.

Levinson (1979, 1989, 1993) advances a related theory for our concept of *art* in general. He proposes that "to be art is, roughly, to be an object connected in a particular manner, in the intention of a maker or profferer, with *preceding* art or art-regards". Specifically, the artist must intend that the object is to be regarded as art objects in the present and past "are or were correctly regarded. . ." (1993, p. 412). This can happen either through the intention that the object be regarded *in specific ways* in which artwork is correctly regarded, such as "with close attention to form", "with awareness of symbolism", and so on (this is what Levinson calls the intrinsical mode of art-making), or through the intention that the object be viewed however art is correctly viewed, without any specific notion as to what this is (this is what Levinson calls the relational mode of art-making).

Under this account, the oddest entities can come to be viewed as artwork.

Levinson (1989) gives the example of Jaspers, who "directs our attention to a pile of wood shavings on the floor, a green 3″ × 5″ index card tacked to his wall, and the fact that Montgomery is the capital of Alabama. He names this set of things *John*. He then says that this is his latest artwork." Levinson argues that we are willing to accept this as an artwork only to the extent that we can be convinced that this collection has been sincerely and correctly intended by Jaspers for regard in a certain way, one associated with how art is typically viewed. This inference is easier for more traditional entities – it would be simpler if Jaspers just put paint onto a canvas – but the history of art is one of an increasingly broader class of entities being viewed as art, and the artwork status of creations such as those by Warhol or Oldenburg (and, before that, the work of Impressionists such as Monet) is by now relatively uncontroversial.

One of the merits of this approach is that it can capture the continuity between what we view as art now and what was taken to be art in the past (as the art of any period is defined in terms of the art prior to it, and thus if something was viewed as art in the past it should still be art now), while at the same time explaining the existence of artistic progress (if the art of a given period can be based on prior art, but not vice versa, then it is not surprising that more things can be art now than in the 1800s). Levinson describes this as an "intentional-historical" theory, as existing art is defined through the intention to create something that (either intrinsically or relationally) is related in the right way to preceding art.[3]

It is worth noting a difference between the theory of pictures above and Levinson's account of art in general. The specific type of representational picture is determined by what the artist intended to *represent*, not how she intends it to be regarded by others, while the status of an entity as art is, according to Levinson, determined by the intention that it should be *regarded* as art. But both views share a crucial assumption, which is that our beliefs about the extension of artifact concepts are based on intuitions about creators' intentions that themselves make reference to members of the specific kinds – a picture of a dog is a picture that is intended by the artist to represent a dog, art is what is intended by the artist to be regarded as art.

4. An intentional-historical theory of artifact concepts

4.1. The proposal

A proposal related to Levinson's can be presented for artifact concepts in general, as follows:

[3] Levinson notes that this theory is incomplete in the sense that it does not explain how the *first* art objects came into existence, and he discusses different ways to expand this notion to account for how such original objects can be counted as art.

> We construe the extension of artifact kind *X* to be those entities that have been successfully created with the intention that they belong to the same kind as current and previous *X*s.

In other words, our understanding of the concept *chair* is that it includes all and only those entities that have been successfully created with the intention that they belong to the same kind as current and previous chairs (or, equivalently, with the intention that they be chairs). The caveat "successfully" expresses the constraint that the entities must turn out as they were intended to turn out. If someone intends to create a chair but it falls to pieces as soon as it is finished, the person would not view this creation as successfully fulfilling his or her intent, and thus has not created a chair. Put differently, we view chairs as those entities that would be viewed by their creators as the successful products of the intention to create chairs.

Note first that this analysis solves the problem that intentions such as "created to be sat upon" are too general and include objects such as sofas and benches; intentions such as "created to be a chair" are exactly specific enough and do not include sofas and benches. And although someone can create a chair without intending anybody to sit on it, it is difficult to see how someone could create a chair without intending it to be a chair.

A few objections to this view can be quickly put aside. First, there is nothing odd about attributing to someone the desire to create a chair. It is surely no odder than attributing to someone to desire to own a chair, or to draw a chair, and so there is no reason to find a theory which posits intentions such as "to create a chair" any less plausible than a theory that posits intentions such as "to create something that you can sit upon". If we grant that intending "to represent a dog" can distinguish pictures of dogs from pictures of cats, then it is at least possible that intending "to create a chair" can distinguish chairs from clocks.

Second, there is nothing tautological about saying that chairs are those entities successfully created with the intention that they be chairs. It could just as well be that it is appearance or current use that makes something a chair, or that this sort of intention is relevant but not sufficient, or that some other sort of intention is more relevant, and many scholars would argue that one or more of these alternatives is correct. For better or worse, then, this is a falsifiable theory.

Finally, it is not, in the normal sense of the term, a *definition* of the word "chair" to say that it refers to those objects that have been successfully created with the intention that they be chairs, as the word "chair" appears on both sides of the equation. But one need not define a word in order to know what it means (a good thing, given that few words appear to have definitions; Fodor, 1981). Even without necessary and sufficient features, we do possess knowledge about chairs, what they typically look like and how they are typically used, and we can use this knowledge – along with our notions about the relationship between intention and product – to infer whether a novel entity was intended to be a chair, just as we can use the same sort of knowledge to infer whether something was intended to be a picture of a dog, or a work of art.

Consider the following scenario, presented in the context of how a person (such as a child, or an adult from another culture) might learn the meaning of the word "chair". She is exposed to her first chair and hears it described as "a chair". On the basis of contextual and linguistic cues, she infers that the word "chair" refers to this object and other objects of the same kind (see Bloom and Kelemen, 1995; Markman, 1990). Based on certain physical properties of the entity, such as the existence of straight lines and right angles, she infers that it is an artifact. (Note that even 9-month-olds can distinguish between toy animals and toy vehicles; Mandler and McDonough, 1993). She notes what the chair looks like and what it is currently used for, and might even infer on the basis of its appearance what function it is likely to have been constructed to fulfill.

After some initial exposure to chairs, what underlies her intuition as to which other objects belong to this kind? Different theories diverge at this point. Definitional proposals (e.g., Katz and Fodor, 1963) posit that she might think that chairs are all and only chair-shaped entities or all and only entities designed for the function of being sat upon. Prototype theories (e.g., Rosch and Mervis, 1975) posit that appearance and function might make different contributions to her judgements as to whether something is a chair. Something shaped like a typical chair and built to sit on is definitely a chair; a beanbag chair is less of a chair, as is a chair that is only used to pile books on; while something of a novel shape that has never been sat upon might not be a chair at all.

The intentional-historical hypothesis posits that the person will assume that chairs are those entities created with the intention to belong to the same kind as the original object or objects she was exposed to. She might note, for instance, that the entities that have been described as chairs tend to share a certain shape, and might observe that people tend to sit on them. If she is then exposed to a novel object of the same shape that is used in the same way, she will assume that this too is a chair, because this sameness of shape and use is an excellent cue that it was intentionally created to be a member of the same kind.

This proposal is different from the position that artifact concepts are individuated solely in terms of the functions that the objects were intended to fulfill, such as sitting upon for chairs. The considerations raised earlier suggest that intended function cannot be criterial, since we are capable of distinguishing chairs from members of other artifact kinds that are also made to be sat upon, such as benches and stools. To tell whether something was created with the intention that it fall into the class of chairs and not the class of benches, then, one must attend to other properties of the object that reflect the creator's intention, such as shape, size, and texture.

This is not to deny, of course, that chairs and benches differ in their typical functions. On the contrary, chairs are standardly created to seat a single person and benches are standardly created to seat more than one person. This difference in typical function explains physical differences between typical chairs and typical benches, such as the fact that benches are usually longer and sturdier than chairs. But such functional properties cannot be critical. If they were, we would not be able to conceive of someone building a large chair to seat several small children, or building a bench with the intention that a single person stretch out on it. But these scenarios are clearly possible. In fact, as noted above, it is consistent with our understanding of artifact kinds that one can build a chair or a bench without intending it to be used at all.

This proposal allows us to posit a procedure through which people determine the specific kind that a novel artifact belongs to, as follows:

> We infer that a novel entity has been successfully created with the intention to be a member of artifact kind *X* – and thus is a member of artifact kind *X* – if its appearance and potential use are best explained as resulting from the intention to create a member of artifact kind *X*.

4.2. Comparison with prototype theory

This proposal should be evaluated in comparison with its most plausible alternative, a prototype theory in which kind membership is directly inferred on the basis of properties such as intended function, current use, and physical appearance. For the example above, the intentional-historical view explains why we would view a chair-shaped object that one could sit on as a chair by positing that we naturally infer that such an object has been successfully created to be a member of the same kind as previously encountered chairs. The prototype theory could easily explain the same fact: this second object is identical in virtually all regards to previously encountered chairs and thus matching to a prototype would lead it to be categorized as a chair.

There is an explanatory difference between these two accounts, however, even for the simple case here. One criticism of prototype theories is that they are unconstrained (e.g., Keil, 1989; Murphy & Medin, 1985); they do not explain why some features are part of a concept and others are not, why some are more important than others, and so on. In the case of artifact concepts, one might posit that features such as current use and intended use (as in "used to sit on" and "created so that people would sit on it") would be weighted highly, and so would the non-intentional feature of being of a certain shape (see also Leyton, 1992). Texture and size would be less important, and color would be irrelevant. This raises the question of *why* this is the case; why are the features ranked in this manner?

The intentional-historical theory offers an explanation. Our intuitions as to which features of an object are relevant when determining artifact kind membership are based on our inferences as to which features exist as the result of the intention to construct a member of a specific artifact kind. In particular, similarities between a novel object and previous members of the kind will be important if one or both of the following conditions hold: First, the similarity is perceived as non-accidental; in the sense that it would be unlikely to occur if it was not intended to occur. We intuit that something is probably not going to be shaped like a typical chair if it was not constructed with the intention to be shaped like a typical chair. In contrast, all sorts of things can be the size or color of a typical chair, and we are thus not driven to an intentional explanation when we observe sameness of size or color. Second, the similarity is perceived as relevant to why the artifact was created in the first place. Given the typical motivation for building chairs, their shape is quite constrained – while, with the exception of street lights and camouflage, the color of an artifact is typically irrelevant. This importance of shape shows up very early in development, both in 2-year-old's strong bias to generalize novel words on the basis of shape (e.g., Landau, 1994), and in their understanding of the correspondences that hold between object shape and object function (McCarrell & Callanan, 1995). In general, then, it is possible that our intuitions as to how and why artifacts are created can provide a framework with which to explain why some properties of artifacts are relevant for categorization and some are not (see also Kelemen & Bloom, 1994).

The prototype theory and the intentional-historical theory also differ in what they predict about cases in which the identity of appearance in the chair example above does not exist. We have no problem viewing a hand-shaped chair as a chair, a duck-shaped clock as a clock, or a green cube as a pawn, and these categorizations cannot be based on properties such as shape. Such examples pose a problem for prototype theories of concepts. Under such theories, something is judged to be a member of a kind to the extent it possesses features that typical members of that kind possess. It is judged to be less of a member to the extent that it lacks these features or has different ones. This type of theory provides a plausible explanation of judgments of typicality, reaction-time responses, and the like (e.g., Rosch & Mervis, 1975), and may explain how we can quickly categorize objects and pictures of objects as belonging to familiar artifact kinds. (It is unlikely that every time a person looks around for a chair, she is drawing inferences about creator's intentions.) Nevertheless it cannot account for our considered intuitions about kind membership. Unusual chairs, clocks, and pawns are not necessarily unclear or fuzzy members of their kinds. On the contrary, there is evidence that people are willing to accept an atypical object as a member of a kind to the extent they can imagine a plausible scenario as to how someone could successfully build it with the intention that it be a member of that kind.[4]

One case of this concerns our intuitions about *historical continuity*, both real history and the present-viewed-as-the-past history. Malt (1991) provides several examples of this. Cardboard frozen orange juice containers are called "cans", most likely because they have evolved from a more prototypical metal version used for the same function. Baby bottles are no longer made out of glass and often diverge in shape from typical bottles, but are viewed as bottles because of their historical development from the more typical bottles that were once used with babies. In some hotels, rooms are opened by a strip of metal or plastic inserted into a slot in the door; this operates an electronic mechanism that opens the lock. These strips are categorized as keys, at least in part because of their historical relationship to more typical keys. In all of these cases, atypical members are assumed to belong to an artifact kind by virtue of our appreciation that they form a natural continuity with previous instances of that kind.

One might object that for at least some of these examples, people who possess the concepts could be unaware of the historical connection; they might view an orange juice container as a can, for instance, simply because they have heard it described as such. Other evidence, however, suggests that intuitions of historical connectedness can directly underlie a person's categorization of artifacts. Recall that Malt and Johnson (1992) found that most of the items described with normal functional descriptions but unusual physical descriptions were judged as not belonging to the artifact kind (as in the example of *boat* presented above). But some *were* consistently judged as artifact kind members despite their unusual physical descriptions. In particular, subjects tended to judge the description "a small cube with an expandable plastic rod that can be placed in contact with a surface to register its distance on a digital display" as consistent with being a ruler; the description "assorted comfortable pillows, a robot that calculates operating costs and collects money, and is painted purple" as consistent with being a taxi; and the description "a large disk suspended from the ceiling by cables, with a fold-down seat attached, and manila folders tacked around the perimeter" as consistent with being a desk.

What distinguishes these categories from the others? As Malt and Johnson point out, their physical features were unlikely to be part of the

[4] The intentional-historical proposal can also explain the existence of unclear or fuzzy cases, but does so in a different way than the prototype theory. Such cases would correspond to instances in which we are unclear what the creator intended (as when we see in a store a clay object intermediate in shape between a cup and a bowl, and have no idea which – if either – it was intended to be), or in which we doubt that a determine intention actually exists (as when looking at the same object created by a very young child who might not herself know the difference between cups and bowls).

subjects' previous experience with members of these kinds. Why, then, were subjects more willing to accept "a small cube with an expandable plastic rod. . ." as a ruler than they were to accept "spherical and made of rubber, is hitched to a team of dolphins. . ." as a boat? Malt and Johnson speculate that "the unusual physical features for these three items may be construed as more advanced or effective than the current features, and the descriptions may be interpreted as *plausible futuristic versions of the articles*" (p. 209; emphasis added).

This speculation is quite consistent with the proposal here. It is not that the "small cube" description shares more features with typical rulers than the "spherical" description shares with typical boats. Instead it is that we can easily imagine the "small cube" item as an object successfully created with the intention to create a better ruler (i.e., a better member of the kind that includes present and past rulers), while it is implausible that someone who sincerely intended to build an object belonging to the class of boats would make the spherical rubber object. In other words, it is not similarity *per se* (in terms of overlap of features or some other non-intentional metric) that governs our intuitions that the cube is a ruler and that the spherical object is not a boat – it is instead our intuitions as to what sorts of entities are likely to emerge out of the desire to create objects that belong to the same kinds as rulers and boats. Such examples might not be all that unusual. It requires no great inferential leap to infer that a straight-backed four-legged medium-sized object that one can sit on is a chair. But in the course of a person's life, she will be exposed to an extraordinary array of chairs, some emerging through technological advance, others that extend the boundaries of fashion or aesthetics, still others the result of fiction, either historical, fantasy, or futuristic. For each, the question is *not* "How similar is this to members of the class of chairs?"; it is "How likely is it that this was successfully constructed with the intention to be a member of the class of chairs?"

The answers to these two questions will overlap, but not entirely, and this suggests a few ways to more systematically compare the prototype theory of artifact concepts with the intentional-historical theory. The prototype theory entails that people should judge a novel entity to be a chair to the extent that it is similar to known chairs, where similarity is captured in terms of some list of features. The intentional-historical account will appeal to other factors as well, such as one's intuitions as to what would constitute an *improvement* on current chairs, in terms of advanced physical structure, enhanced function, or just better style. Based on the observations of Malt and Johnson, for instance, one would predict that given two objects judged to be equally dissimilar to existing chairs (and thus equivalent from the standpoint of prototype theory), subjects will judge the object that is more useful, or more futuristic, to be more likely to be a chair. One should also expect to find "chaining", such that if subjects believe object X to be a futuristic version of a current chair, then they should judge that object Y, which is a futuristic version of object X, to also be a chair, even if Y is highly dissimilar from known chairs.

Another difference between the two views is that the intentional-historical theory predicts that the known intentions and the past activities of the creator of an artifact should be relevant to determining the kind that this artifact belongs to, even if they don't have anything directly to do with its appearance or use. Given an object that is intermediate in appearance between a chair and a bench, for instance, one would expect people to be sensitive to what the creator of the object claims to have intended it to be, or what kinds of artifacts she has built in the past. As a very minimal case, subjects should give a lot of weight as to whether the creator calls the object "a chair" or "a bench", and this should matter more than what the person who buys the object calls it, or even what an expert the history or construction of chairs and benches calls it. This is clearly our intuition for representational pictures. Given a picture that is intermediate in appearance between two objects, such as one that looks equally like a fork and a spoon, the adult intuition is that the identity of the picture is determined by what the artist intended it to be a picture of, as indicated by how the artist describes the picture or what she was looking at while she drew it (and this is also the judgment of 4-year-old children; Bloom & Markson, in preparation). But the intuitions are perhaps less clear for artifact kinds in general, even for adults, and these issues remain to be addressed in future research.

4.3. Destruction and transformation

A related issue concerns the nature of our intuitions about destruction and transformation. What factors do we see as causing an object to cease being a member of an artifact kind, as when a clock is smashed to dust and is no longer a clock? And what are our intuitions about what can cause an object to *become* a member of an artifact kind without being created in a literal sense, as when a penny becomes a pawn or a piece of furniture becomes an artwork?

A preliminary hypothesis as to our intuitions about when an object ceases to be a member of an artifact kind is as follows:

> A damaged member of an artifact kind X will be construed as remaining a member of artifact kind X only to the extent that it can be *fixed* – to the extent that it can, as the result of further intentional action, return to be a functioning member of artifact kind X.

Consider a clock that does not tell time. Imagine first that it does not work because it needs to be rewound, or because its batteries have died. It is clearly still a clock, perhaps because fixing it (rewinding it, replacing the batteries) is quite simple. If one breaks it by hitting it (gently) with a hammer, one can still bring it back to working order, and it is still a clock, but the more damage one does, the harder it is to repair, and we are increasingly more reluctant to view it as a clock. Taken to the extreme, when one smashes the clock to dust, it is no longer a clock at all, because it can never be fixed.

Other considerations, however, suggest that "fixability" does not adequately characterize our intuitions about artifact kinds. Imagine that our clock was broken because a small internal piece wore out, but that it could never be repaired, as the piece could not be replaced. Still, it would remain a clock. A different example, adopted from Hirsch (1982), is as follows. Imagine taking apart the clock on Monday, separating it into a hundred pieces and sending off the pieces to different parts of a factory to be cleaned. On Wednesday, the pieces of the clock are returned and you reassemble the clock. It is clearly the same clock as it was on Monday. But it is also our intuition that this clock ceased to exist on Tuesday – one would not naturally view the set of distinct parts scattered across the factory on Tuesday as a single clock. This is despite the fact that this set of parts could be (and will be) reassembled into a functioning clock on Wednesday. These considerations suggest that "fixability" is neither necessary nor sufficient for a damaged artifact to remain a member of an artifact kind.[5]

A more plausible account is as follows (this is an expansion of a proposal made above):

> We infer that a novel entity has been successfully created with the intention to be a member of artifact kind X – and thus is a member of artifact kind X – if its *current* appearance and potential use are best explained as resulting from the intention to create a member of artifact kind X.

A clock that needs to be rewound or that does not work because it has been gently hit by a hammer is still viewed as a clock, because the details of its structure are such that the best explanation for how it came into existence is through the intention to create a clock. This is the case even if it can never be repaired. A pile of dust created by hitting a clock very hard with a hammer is not judged to be a clock since its current appearance and use are not best explained in terms of the desire to create a clock. Note that the question is not whether one is *aware* of the intention (by that standard, it would be impossible for something that one knows to have been created to be a clock to ever stop being a clock, which is plainly not our intuition); it is whether the *current* status of the entity is consistent with this intention.

A final issue concerns transformations, where something that already

[5] Similar considerations apply for our intuitions about non-artifact kinds as well. A brain that is severely and irreparably damaged is still a brain, for instance. It is conceivable that a suitably modified theory could also explain our intuitive beliefs about biological organs, since, as discussed below, these are conceptualized in certain regards very much like artifacts.

exists as a member of one artifact kind becomes a member of another artifact kind. Sometimes this can occur through physical manipulation, and thus can be explained under the same theory that applies to artifacts in general. If one drops and cracks a coffee pot, it is still a coffee pot, but if one carefully reshapes it, adds and removes parts, so that it can feed birds and looks like a bird feeder, both adults and children would say that it is no longer a coffee pot, it is a bird feeder (Keil, 1989). Similarly, houses can become churches, computer monitors can become fish tanks, and swords can become plowshares.

A particularly interesting case is when novel artifacts are created solely as a result of subsequent intention. Pennies can become pawns, for instance, and baseball bats can become weapons. In some of these instances, the object retains its original categorization; it is a penny and a pawn, a bat and a weapon. What these cases suggest is that there are more ways to "create" something than to build it out of raw material. Many artifacts *do* need to be made from scratch given the sorts of functions they typically fulfill; this is especially the case for members of basic-levels kinds with typically distinctive shapes such as chairs and for members of kinds that fulfill functions that are especially complicated from an engineering perspective, such as VCRs. But certain artifacts can be created out of members of different artifact kinds or even from natural kinds. Such cases include kinds like *weapon*, *pawn*, *toy*, *paperweight*, *target*, *landmark*, *coin*, and *artwork*.

Given that the creation of members of these kinds need not involve any physical changes, our judgments are driven here solely by our intuitions about intentionality. We construe a penny as a pawn only if we intuit that a person sincerely intends for the penny to fall into the class of pawns. The right sort of intention is essential; someone who does not know the rules of chess could not do this, as she lacks the right understanding of what it is for something to be a pawn. Note that one does not have to *do* anything to the penny for it to become a pawn. One can put a penny in the middle of a cluttered chess set and start to think about a chess problem – what makes this penny a pawn (as opposed to a queen, say) is the mental state of the person who is considering the chess problem (for discussion of a similar case, see Haugeland, 1993).

In the domain of art, this process is sometimes called "transfiguration" (Danto, 1981). Duchamp's *Fountain* was originally a urinal; his *In Advance of a Broken Arm* was once a mere snow shovel. Museums are full of religious and functional artifacts that are now viewed as "art" even though many were originally intended to fall into very different kinds. What these instances share with the pawn case above is that the acts of transfiguration (from snow shovel into sculpture, for instance) are the result of intentional acts, not physical ones.

Several issues arise here that go beyond the scope of this paper. One is developmental: can young children appreciate these sorts of abstract kinds? There is some evidence that they can; consider their early acquisition of expression like "toy", where toys are objects intended to be toys, and do not have to be specifically constructed for this purpose (see also Wooley & Wellman, 1990). A further issue concerns our intuitions about the conditions under which something (or someone) can be made into a member of an artifact kind, such as an artwork. As Levinson (1993, p. 418) puts it: "Transfiguration is a serious business. . . and strange as it seems, there must be moral and legal limits to it." The theory presented here has nothing to say about the nature of such limits.

To sum up so far, it was argued that we decide whether something is a member of an artifact kind through our intuitions as to whether it was successfully created with the intention that it belong to that kind. This was contrasted with a prototype theory in which kind membership is determined by comparing the features of an object to the features that previous members of that kind possess. Several considerations favor the intentional-historical proposal. First, the fact that a feature such as shape (but not color) is highly important for determining kind membership must be stipulated under a prototype view, but can be explained by the intentional-historical account in terms of our intuitions that identity of a feature such as shape (but not color) is both unlikely to occur by accident and relevant to why most artifacts are intentionally brought into existence. Second, we infer that objects with features that are different from those of previously experienced members of a kind can belong to that kind if these objects can be historically related to other members of the kind, either as current developments from past instances (as in a cardboard container viewed as a can; Malt, 1991) or as futuristic developments from current instances (as in a robot taxi; Malt & Johnson, 1992). This follows naturally from the intentional-historical proposal, but cannot be explained by any account that proposes that we determine kind membership solely on the basis of similarity. Finally, our intuitions about persistence and transformation also appear to be based the perceived intention of the creator: We infer that broken and unreparable objects can remain members of an artifact kind so long as they are still identifiable as being created with the intention to belong to that kind (as with a broken clock) and that an object can become a member of an artifact kind solely through the intention of the creator (as when a penny becomes a pawn).

4.4. Objections and modifications

Consider now a couple of criticisms that could be raised against the intentional-historical proposal. The first concerns cases in which intention does not appear to be *sufficient* for us to view something to be a member of an artifact kind; the second concerns cases in which it does not appear to be *necessary*.

Imagine a madman who creates a tiny pile of dirt, assuming that people will happily sit on it, and he states that this pile was successfully created with the intention to be a chair. Still, we would not view it as a chair. We would have a parallel response to a brain-damaged artist who carefully draws something that looks exactly like a cat, and proudly describes it as a picture of a dog. Or to a 2-year-old who creates a flat disk out of clay and claims that it is a cup. Finally, consider *Deadman*, created in 1972, in which Chris Burden had himself enclosed in a sack and placed on a California freeway. Burden (as well as Danto, 1986) clearly viewed this as a work of art, but many would disagree – and this is not because we doubt the sincerity of someone who places himself in the middle of traffic in order to make an aesthetic point. These examples can be taken to suggest that even if we believe that an object was created with the sincere intent to be a member of certain kind, this is not always enough for us to believe that the object is in fact a member of the kind.

All of these examples share a certain property, however. When the madman describes a pile of dust as the successful result of an attempt to create a chair, it is clear that his understanding of chairs is quite different from our own, perhaps so much so that it is not actually correct to say that he is in an intentional state that makes reference to the same concepts that we possess. Similarly, we are likely to infer that the child who calls a disk "a cup" does not really know what a cup is, or is perhaps confusing the words "cup" and "plate". We would respond in the same manner to the unfortunate who thinks she drew a dog when she actually drew a cat. Finally, those people who doubt that Burden's creation counts as "art" are likely to differ from him as to what *other* entities they view as art. In general, intentional attribution might require significant conceptual overlap – we attribute to someone in the intention "to create *X*" under the assumption that their experience with *X*s and their beliefs about the nature of *X*s are to a large extent consistent with ours (for different perspectives on this issue, see Dennett, 1987; Fodor & LePore, 1992). In the cases above, this requirement is arguably not met; and thus we do *not* actually attribute to these people the intentions to create a chair, a cup, a picture of a dog, or (for some of us, at least) a work of art.

This leads to the prediction, which seems correct, that there will be no cases in which someone who we view as having a *similar* conceptual scheme to our own creates something that she views as a member of artifact kind *X* but which is really, by our own judgment, not a member of artifact kind *X*. The cases that do exist will involve instances in which the conceptual schemes of others are understood on independent grounds as being quite different from our own, as when artifacts are created by the mad, by the brain-damaged, by children, or by those who occupy the cutting-edge of art and philosophy.

The second sort of criticism is often discussed in the philosophy literature, and goes as follows: Suppose a lighting bolt hit a rock and, through some wildly improbable coincidence, transformed the rock into an object of exactly the same size and shape of a typical desk chair. Under the intentional-historical theory defended here, people should not view it as a chair, as it was not created with any intention at all. Intuitions differ. Some people (mostly philosophers) have the expected intuition – it looks like a chair, could be used as a chair, but it is not, the instant after the lighting strikes, a chair. Others, however, would view this object as a chair (and not merely a potential chair). This sort of intuition refutes the claim that members of specific artifact kinds are necessarily seen as emerging through a specific sort of intent.

This is a genuine counter-example to the intentional-historical theory, as it shows that intention is not criterial. Examples can be multiplied. If a can of paint spills onto a canvas and forms something that looks exactly like a picture of a cat, it is likely that we would view it as a picture of a cat. We might respond in the same way to the creation of the brain-damaged artist discussed above, insisting that she drew a picture of a cat even if we believe that she is sincere in denying that this was what she planned to do.

There are two ways to expand the theory above so as to account for these cases. One is to allow that people *can* be swayed solely by superficial features, independently from notions of intention – we view the rock-chair as a chair solely because it has the same appearance and potential use as a typical chair. A different alternative, more amenable to the spirit of the intentional-historical proposal, is to posit that something can be a member of an artifact kind if it is of the physical structure that would arise *if* it was created with the intention to be a member of an artifact kind.

These are two reasons to favor this second analysis. First, it explains why our intuitions are swayed by the perceived complexity and non-randomness of the object we are categorizing. We are more prone to categorize the rock as a chair if it looks like a highly structured artifact with many different parts than if it looks like a simple artifact with only a few parts. This is not because a complex chair is necessarily a more typical chair than a simple chair. Instead, it is perhaps because of our intuition that the more complex an object is, the more likely it is to have been created through an intentional process. To the extent that there is an effect of perceived complexity on whether we categorize non-intentionally created objects as members of artifact kinds, this would suggest that such categorizations are the result of viewing some objects "as if" they were intentionally created – and are not directly due to sameness of appearance and potential use.

A second consideration is that for the exceptional examples above, there exists a considerable pull towards the intuition that there actually is an intentional process at work. This is reflected in the well-known "Argument from Design", in which the adaptively complex design that pervades the natural world is taken as evidence for some intentional creator, usually God (e.g., Paley, 1802). The neo-Darwinian theory of evolution has provided a scientific explanation for the creation of complex design without an intentional designer (Darwin, 1859; see also Dawkins, 1986), but many people are nevertheless drawn towards the intentional account, both adults and young children (Keleman, 1995). Just as the complex structure of the eye led theologians to infer that it was an artifact created by a benevolent God, the highly intentional appearance of the rock-chair might license the inference that it too was created (or is of the same form *as if* it was created) by some intentional force that is basing its actions on knowledge of current and previous chairs. And this might lead us to categorize it as actually being a chair.[6]

The intentional-historical theory posits that people should infer a strong relationship between kinds of artifacts and kinds of intentional states, and the responses to both of the above concerns are consistent with this position: Those instances in which there is an apparent intention to create an X without an X having actually been created (as when a madman views a pile of dust as the successful outcome of the intention to create a chair) are restricted to cases in which the conceptual schemes of the creators are so different from our own that there is independent motivation to doubt that the right intentions actually exist. And those unlikely instances in which an X is created without any associated intention (as with a lighting bolt transforming a rock into a chair) are of the same sort that have historically led to the intuition that some non-human intentional entity was at work in the creation of the natural world.

5. Implications for theories of concepts

How does the theory outlined above for artifact kinds relate to more general proposals about the nature of concepts? The average English-speaking adult knows over 60,000 words (Pinker, 1994) and for each she has some notion as to what the word does and does not correspond to. Traditional theories of concepts posit that this knowledge (the "intension" of the word) is encoded as a set of features (or properties) that correspond in some manner to the features (or properties) of the entity or entities that the word refers to. Definitional theories posit that these features provide necessary and sufficient conditions for kind membership, while prototype theories posit that whether something is judged to belong to a kind is a relative matter, determined by the extent to which the object shares the same features as typical members of that kind.

This decompositional program has not been particularly successful, however. Psychologists, philosophers, linguists, and anthropologists have

[6] In a limited sense, this modified proposal is still consistent with the procedure for determining artifact kind membership introduced above (". . . if its current appearance and potential use are best explained as resulting from the intention to create a member of artifact kind X"), since the current appearance and potential use of entities like the rock-chair *are* best explained in terms of intention – it is just that the best (most plausible, least mysterious) explanation is not the one that is actually correct.

found few plausible decompositions for words in natural language. Most people do not appear to encode a set of necessary and sufficient features for kinds like *chair* and *lemon* (Fodor, 1981; Smith & Medin, 1981). Similarly, there is little evidence that there exist clusters of features that capture our (presumably graded) intuitions as to what entities are chairs and lemons. Although prototype theories provide some account for our intuitions about typicality, and might even explain how people rapidly categorize visually presented objects, they fail to explain our considered intuitions about kind membership (Armstrong *et al.*, 1983; Landau, 1982; Margolis, 1994).

This failure to find such exhaustive decompositions appear to be for principled reasons (Kripke, 1980; Putnam, 1975, 1977). We might associate certain features with concepts like *lemon* (Putnam) or *Gödel* (Kripke), but our intuition is that objects can be in the extensions of these concepts even if they lack these features, and that objects can possess these features and not be in the extensions of the concepts. Lemons are viewed as typically having the features of a yellow peel and tart taste, but we could accept an abnormal lemon that is blue and sweet; Gödel is associated with the property of having generated a certain proof in mathematics, but we could well imagine that we are mistaken and that the proof was done by someone else. Similarly, not every yellow-peeled tart-tasting entity is necessarily a lemon and not anyone who generated this mathematical proof is necessarily Gödel.

These considerations have led to a shift in perspective as to the nature of concepts, particularly for natural kinds, within cognitive and developmental psychology (e.g., Carey, 1986, 1988; Gelman, 1988; Gelman & Markman, 1986, 1987; Keil, 1989; Malt, 1994; Medin & Ortony, 1989; Murphy & Medin, 1985). Although we possess an understanding of typical features of kinds and individuals, we also know these features are not criterial. Under one version of this theory, our concept *lemon* is "those entities that belong to the same kind (have the same essence) as previously encountered lemons". Contemporary adults believe that this "essence" is expressed through some genetic theory for biological kinds and some atomic theory for chemical kinds, but a belief in essences exists across cultures, expressed in quite different ways, and is also present in young children (e.g., Gelman, 1988; Keil, 1989).[7]

This essentialist theory is not usually extended to artifacts, however (see Abbott, 1989; Schwartz, 1977). As the point is usually made, one might study dogs to find out what deep essence they share, but one could not normally study chairs to find out what *really* makes them chairs. In contrast, the proposal here is that the mental representation of artifact kinds is quite similar in structure to that of natural kinds. Just as our notion of *lemon* is

[7] This fact about human psychology does not imply that essences actually exist, of course. For instance, Atran (1990) notes that people persist in believing that "tree" is a biological natural kind despite the fact that botanists say otherwise (see also Dupre, 1981; Hull, 1965; Lakoff, 1987; Malt, 1991).

"those entities that belong to the same kind as previously encountered lemons", our notion of *chair* is "those entities that have been successfully created with the intention to belong to the same kind as previously encountered chairs". In both cases, we use superficial features such as appearance to infer what entities are likely to "belong to the same kind" or to "have been successfully created with the intention to belong to the same kind", but in both cases, the relationship between these features and kind membership is indirect, and is mediated through a naive theory. Thus there are circumstances under which something can be a lemon or a chair even if it lacks the typical features of lemons or chairs and something might not be a lemon or a chair even if it does possess the typical features of members of those kinds.

One implication of this is that many expressions within natural language – including proper names, natural kind terms, and artifact terms – correspond to concepts that do not exhaustively decompose into simpler notions. This should not be surprising. Consider again the traditional example of proper names. When we hear that someone is named "Gödel", we do not assume that this name corresponds to his appearance or to the tone of his voice or to what he is described as having accomplished. Instead we assume that the word corresponds to *that person*, and we encode this other information as contingent properties of that person. In this regard, it is possible for us to question whether Elvis is really dead, whether Shakespeare wrote *Hamlet*, *Macbeth*, etc. (even if for some of us just about all we know about Shakespeare is that he wrote *Hamlet*, *Macbeth*, etc.), whether Moses really crossed the Red Sea, and so on (Kripke, 1980).

This is a plausible way for our conceptual system to work; we want the capacity to think about, recognize, and track *people*, not properties of people. Similarly, we need the capacity to think about kinds that occur in the physical and social world, kinds like chairs and threats and tigers and lemons, to reason about them, have desires that implicate them, and so on – but to also understand that things can be members of these kinds even if they appear different from typical members, or that they might not be members of these kinds even if they appear similar to typical members.

Many cognitive scientists find the notion that concepts such as *chair* and *Gödel* are not exhaustively reducible to a set of features to be an extreme doctrine, but this might be due in part to a misunderstanding of what it implies about human psychology. (It does not entail, for instance, that concepts such as *chair* and *Gödel* need to be innate.) What it does entail is the following.

We possess the capacity to create novel cognitive entries when acquiring new concepts, corresponding either to individuals (as with *Gödel*), artifact kinds (as with *chair*), or natural kinds (as with *lemon*). We might be motivated to create such an entry by hearing a novel word, and cues such as the syntax of the word and the discourse context in which it appears can aid a person in determining which type of category the entry belongs to (Bloom, 1994; Hall, 1994; Tomasello & Barton, 1994). This entry would include known properties of the members, both those that we observe (such as the shape of a typical chair) and those that we infer through linguistic and non-linguistic context (such that Gödel was a mathematician). This knowledge is accumulated through experience with members of the kind and is likely to be encoded as a prototype structure.

The considerations raised above in the case of artifacts, and the more general arguments by philosophers like Kripke and Putnam for individuals and natural kinds, suggest that this is not sufficient as a theory of concepts. It constitutes our representation of previous members of the kind, which is just one aspect of how we infer the extension of the concept. We also require a rich set of inferential capacities; these include:

- Intuitions about the identity conditions of individuals over time and space. This enables us to determine the conditions under which entity A at time t and location x is the same individual as entity B at time $t + 1$ and location y (for discussions, see Hirsch, 1982; Macnamara, 1986; Spelke, 1994; Wiggins, 1980; Xu & Carey, in press). This is essential for understanding concepts like *Gödel*. It is also a prerequisite for understanding concepts that correspond to kinds of individuals, such as *chair* and *lemon*; one cannot determine what entities fall into these kinds without the prior ability to track and individuate these entities. Note that it is unlikely that a single procedure underlies our intuitions about identity for all types of individuals. More plausibly, our intuitions about the identity conditions for entities like chairs are going to be quite different from our intuitions about entities like people, parties, and countries (see also Chomsky, 1992).
- Intuitions about the "essence" of natural kinds, how this essence is typically reflected in the surface features of members of these kinds, and what transformations on these entities will cause them to no longer belong to their kind. Some of this knowledge might be innate, such as the intuition that such essences exist at all for certain kinds, while other aspects of this knowledge are acquired through cultural experience and scientific training (for discussion, see Carey, 1986. 1988, Keil, 1989; Gelman, 1988; Gelman & Markman, 1986, 1987; Malt, 1994). This is essential for our understanding of concepts like *lemon*.[8]
- Intuitions about how the intentions of people are expressed in the

[8] The distinction between natural kinds and other kinds might not be as sharp as assumed here. Malt (1994) finds that there is a sense in which water is *not* viewed as H_2O, such that, for instance, tea is not categorized as a kind of water but ocean water is, despite the fact that we judge tea to have a higher proportion of H_2O than ocean water (see also Chomsky, 1992). This suggests that there is a notion of *water* that is more like an artifact kind, one that is distinct from tea because of tea's distinct status as a beverage. Nevertheless, as Malt herself notes, there does appear to be a notion of water corresponding to a more essentialist H_2O notion; this shows up in utterances such as "Tea is 90% water" or "People are mostly made out of water".

properties of the artifacts that they create. This has been the topic of much of the above discussion (see also Keil, 1989; Malt, 1991; Malt & Johnson, 1992; Rips, 1989). This is essential for our understanding of concepts like *chair*.

This paper began with the question of what underlies our intuitions as to which things in the world are chairs, clocks, and pawns. The theory proposed here is that our intuitions are based on what things in the world we infer have been successfully created with the intention that they be chairs, clocks, and pawns. More generally, we view something as a member of a specific artifact kind if its current appearance and potential use are best explained as resulting from the intention to create a member of that artifact kind. This theory raises several questions. What assumptions underlie our understanding of the relationship between entities in the world and the intentions of the people who create them? Are there developmental or cross-cultural differences in how people construe this relationship? Can the gradual "evolution" of artifacts (such as chairs: Joy, 1967) over human history be explained through this sort of intentional-historical account? And how does an intention-based understanding of artifact kinds relate to teleological (Paley, 1802) and adaptations (Darwin, 1859) construals of the origin of complex biological structures? To the extent that an intentional-historical view can explain phenomena not captured by more traditional accounts, these questions become well worth pursuing.

Acknowledgements

I am very grateful to the following people for comments on earlier versions of this paper: Thomas Bontly, Amy Fountain, Deborah Kelemen, Barbara Landau, Jerold Levinson, Lori Markson, Edward Stein, two anonymous reviewers, and especially Bill Ittelson and Karen Wynn. Preparation of this manuscript was supported by grants from the Spencer Foundation and the Sloan Foundation.

References

Abbott, B. (1989). Nondescriptionality and natural kind terms. *Linguistics and Philosophy, 12*, 269–291.

Armstrong, S., Gleitman, L., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263–308.

Atran, S. (1990). *Cognitive foundations of natural history*. Cambridge, UK: Cambridge University Press.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115–147.

Bloom, P. (1994). Possible names. the role of syntax–semantics mappings in the acquisition of nominals. *Lingua*, *92*, 297–329.

Bloom, P. (in press). Theories of word learning: rationalist alternatives to associationism. In T.K. Bhatia & W.C. Ritchie (Eds.), *Handbook of language acquisition*. New York: Academic Press.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition*, *56*, 1–30.

Carey, S. (1986). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language*, *3*, 167–181.

Chomsky, N. (1992). Explaining language use. *Philosophical Topics*, *20*, 205–231.

Danto, A. (1981). *The transfiguration of the commonplace*. Cambridge, MA: Harvard University Press.

Danto, A. (1986). Art and disturbation. In A.C. Danto (Ed.), *The philosophical disenfranchisement of art*. New York: Columbia University Press.

Danto, A. (1992). The Art World revisited: comedies of similarity. In A.C. Danto (Ed.), *Beyond the Brillo Box: The visual arts in post-historical perspective*. New York: Farrar, Straus, Giroux.

Darwin, C. (1859). *The origin of species*. London: John Murray.

Davies, S. (1991). *Definitions of art*. Ithaca, NY: Cornell University Press.

Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dupre, J. (1981). Natural kinds and biological taxa. *Philosophical Review*, *90*, 66–90.

Fodor, J.A. (1981). The current status of the innateness controversy. In J.A. Fodor (Ed.), *Representations*. Cambridge, MA: MIT Press.

Fodor, J.A., & LePore, E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell.

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

Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–208.

Gelman, S.A., & Markman, E.M. (1987). Young children's induction from natural kinds: the role of categories and appearances. *Child Development*, *58*, 1532–1541.

Goodman, N. (1976). *Languages of art*. Indianapolis: Hackett.

Hagen, M.A. (1986). *Varieties of realism: Geometries of representational art*. Cambridge, UK: Cambridge University Press.

Hall, D.G. (1994). How children learn common nouns and proper names. In J. Macnamara & G. Reyes (Eds.), *The logical foundations of cognition*. Oxford: Oxford University Press.

Hall, D.G. (1995). *Artifacts and origins*. Unpublished manuscript, Department of Psychology, University of British Columbia.

Haugeland, J. (1993) Pattern and being. In M. Rollins (Ed.), *Danto and his critics*. Cambridge, MA: Blackwell.

Hirsch, E. (1982). *The concept of identity*. New York: Oxford University Press.

Hochberg, J., & Brooks,V. (1962). Pictorial recognition as an unlearned ability: a study of one child's performance. *American Journal of Psychology*, *73*, 624–628.

Hull, D.L. (1965). The effect of essentialism on taxonomy: two thousand years of statis. *British Journal for the Philosophy of Science*, *15*, 314–326.

Ittelson, W. (in press). The visual perception of markings. *Psychonomic Bulletin and Review*.

Jackendoff, R. (1992). Mme Tussaud meets the binding theory. *Natural Language and Linguistic Theory*, *10*, 1–31.

Joy, E.T. (1967). *Chairs*. New York: Country Life Books.

Katz, J.J., & Fodor, J.A. (1963). The structure of a semantic theory. *Language*, *39*, 190–210.

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

Kelemen, D. (1995). *The nature and development of the teleological stance*. Poster presented at the biennial meeting of the Society for Research in Child Development, April, 1995.

Kelemen, D., & Bloom P. (1994). Domain-specific knowledge in simple categorization tasks. *Psychonomic Bulletin and Review*, *1*, 390–395.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Labov, (1973). The boundaries of words and their meanings. In R. Shuy & C.S. Bailey (Eds.), *New ways of analyzing variation in English*. Washington, DC: Georgetown University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: Chicago University Press.

Landau, B. (1982). Will the real grandmother please stand up? *Journal of Psycholinguistic Research*, *11*, 47–62.

Landau, B. (1994). Where's what and what's where: the language of objects in space. *Lingua*, *92*, 259–296.

Landau, B., Smith, L.B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Levinson, J. (1979). Defining art historically. *British Journal of Aesthetics*, *19*, 232–250.

Levinson, J. (1985). Titles. *Journals of Aesthetics and Art Criticism*, *44*, 29–39.

Levinson, J. (1989). Refining art historically. *Journal of Aesthetics and Art Criticism*, *47*, 21–33.

Levinson, J. (1993). Extending art historically. *Journal of Aesthetics and Art Criticism*, *51*, 411–423.

Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.

Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.

Malt, B.C. (1991). Word meaning and word use. In P. Schwanenflugel (Ed.), *The psychology of word meanings*. Hillsdale, NJ: Erlbaum.

Malt, B.C. (1994). Water is not H_2O. *Cognitive Psychology*, *27*, 41–70.

Malt, B.C., & Johnson, E.C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, *31*, 195–217.

Mandler, J.M., McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*, 291–318.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, *51*, 73–89.

Markman, E.M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.

McCarrell, N.S., & Callanan, M.A. (1995). Form–function correspondences in children's inference. *Child Development*, *66*, 532–546.

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

Millikan, R.G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.

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

Paley, W. (1802/1851). *Natural theology: Evidence of the existence and attributes of the deity, collected from the appearances of nature*. Boston: Gould & Lincoln.

Pinker, S. (1994). *The language instinct*. New York: Norton.

Putnam, H. (1975). The meaning of "meaning". In H. Putnam (Ed.), *Mind, language, and reality: Philosophical papers* (Vol. 2). Cambridge, UK: Cambridge University Press.

Putnam, H. (1977). Meaning and reference. In S.P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press.

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

Rosch, E., & Mervis, C.B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Schwartz, S.P. (1977). Introduction. In S.P. Schwartz (Ed.), *Naming, necessity, and natural kinds*. Ithaca, NY: Cornell University Press.

Smith, E.E., & Medin, D.L. (1981). *Categories and concepts*. Cambridge, MA: MIT Press.

Soja, N.N., Carey, S., & Spelke, E.S. (1992). Perception, ontology, and word meaning. *Cognition*, *45*, 101–107.

Spelke, E.S. (1994). Initial knowledge: six suggestions. *Cognition*, *50*, 431–445.

Tomasello, M., & Barton, M. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*, 639–650.

Wiggins, D. (1980). *Sameness and substance*. Oxford: Basil Blackwell.

Woolley, J.D., & Wellman, H.M. (1990). Young children's understanding of realities, nonrealities, and appearances. *Child Development*, *61*, 946–961.

Xu, F., & Carey, S. (in press). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*.