The real deal: What judgments of *really* reveal about how people think about artifacts

Barbara C. Malt · Michael R. Paquet

Abstract It is widely assumed that artifacts fall into distinct kinds. These kinds are generally identified by appeal to words—*chair* versus *stool* versus *bowl* versus *vase*, and so on. But contextual and cross-linguistic variation in what artifacts are grouped together by name raise questions about whether artifacts indeed do fall into fixed kinds. Can judgments of what artifacts *really* are reveal a true kind membership, distinct from what the objects are called in communicative contexts? In two experiments, we examined what drives judgments of what an artifact *really* is and what these judgments can tell us about how people think about artifacts. In both experiments, we found that people failed to treat artifacts as having a definitive kind membership in their judgments of what the artifacts *really* were. Instead, *really* judgments reflected the typicality of objects with respect to the things normally called by the queried name. If these judgments are taken as direct evidence about the existence of artifact kinds, the outcome argues against such kinds. Alternatively, *really* judgments themselves may be fundamentally linguistic in nature, and so unable to tap into underlying kind memberships. In either case, if such kinds exist, they remain to be found. A more likely reality may be that intuitions about the existence of artifact kinds reflect the partial clustering of objects in similarity space, plus the fact that each language provides names for some constellations of objects in that space.

Keywords Categorization · Concepts · Artifacts · Kinds

The notion of kinds pervades much of the literature on concepts and conceptual development in psychology and philosophy of mind. The meaning of *kind* is most clear-cut when the term is applied to entities created by nature. Here, there is broad agreement that kinds have an objective reality, given by the distribution of properties in nature. In its most general sense, the term *natural kinds* refers to groupings of entities that share many deep as well as superficial properties, such as the groupings of plants or animals that form scientific genera. The essentialist perspective holds that, more specifically, natural kinds are groupings determined by the presence of a single shared underlying essence, whether or not that essence has been identified by humans (e.g., Putnam, 1975).

Although artifacts are created through deliberate acts by humans or other agents, it is commonly assumed that artifacts also fall into distinct kinds (e.g., Bloom, 1996; Dewar & Xu, 2009; Elder, 2007; Futo, Téglás, Cisbra, & Gergely, 2010; Grandy, 2007; Hauser, 1997; Keil, 1989; Phillips & Santos, 2007; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Thomasson, 2007). Again, the more general sense of the term is simply groupings of entities sharing many properties. Parallel to the case of natural kinds, some theorists have further argued that membership in artifact kinds is determined by the presence of an essence—either the creator's intended function (e.g., German & Johnson, 2002; Kelemen & Carey, 2007; Kemler Nelson, Herron, & Morris, 2002; Matan & Carey, 2001; Putnam, 1975) or intended kind membership (Bloom, 1996, 2000).

Whereas claims for the existence of natural kinds can draw on morphological and genetic evidence from biology as well as on evidence from everyday language (the labeled distinctions of *dog* versus *cat* versus *horse* versus *zebra*, etc.), artifact kinds are generally identified only by appeal to words. That is, it is assumed that kinds correspond to labels such *chair*, *stool*, *bowl*, and *vase*. But the flexibility of artifact naming within and across languages creates some challenges for the notion of artifact kinds. Within a language, an object can be called by contrasting basic-level

B. C. Malt (✉) · M. R. Paquet
Department of Psychology, 17 Memorial Drive East,
Lehigh University,
Bethlehem, PA 18015, USA
e-mail: barbara.malt@lehigh.edu

names—*piano stool* at one time and *end table* at another, *bottle* at one time and *vase* at another, and so on—depending on the current use (e.g., Labov, 1973; Malt & Sloman, 2007). This observation seems problematic for the essentialist possibility of artifact kinds defined by the creator's intended function or intended kind membership, because it suggests that an object's kind can change on the basis of the current use, regardless of its creator's original intention. Variation across languages raises questions about the more general notion of kinds as formed around naturally occurring property clusters. Languages can have substantially different patterns of grouping artifacts by name (see, e.g., Ameel, Storms, Malt, & Sloman, 2005; Kronenfeld, Armstrong, & Wilmoth, 1985; Malt, Sloman, Gennari, Shi, & Wang, 1999). For instance, English separates seating for one person (*chair*) from seating for several people (*sofa*), but Chinese separates hard, non-upholstered seating (*yizi*) from padded, upholstered seating (*safa*), regardless of the number of people seated. Objects called *bottle* in English are not all *botella* in Spanish (Malt et al., 1999), and the objects called *ball* in English are not all *bola* in Spanish, *balle* in French, *bol* in Dutch, and so on. These differences suggest that self-evident groupings of artifacts on the basis of shared properties may not exist.

An alternative to giving up the notion of independently existing artifact kinds on the basis of such evidence is to consider that ordinary naming practices may not be the best way to identify such kinds. In light of the flexibility of naming and its fundamentally pragmatic nature—to achieve communication under situational constraints—one could argue that there might be something that an object *really* is that is not always well reflected in the name that it is given in everyday language use. Intuitions support a distinction between what something is called or looks like and what it really is. Most people would probably agree that objects called *toy gun* are not really guns, for instance, and ones called *shoe tree* are not really trees (see also Bloom, 2007). Questions about what something *really* is have been used in the past to tap into underlying perceptions that may differ from superficial appearance (e.g., Keil, 1989; Taylor & Flavell, 1984). In fact, the very notion of artifact "categorization" as a nonlinguistic process (e.g., Bloom, 2007; Rosch, 1978) implies that people have an understanding of object kindhood that can be independent of their name choices. Thus, judgments about what an artifact *really* is, rather than about what it may be called in some conversational context, may be a better indication of underlying kind membership and may provide more clear-cut evidence for the existence of kinds.

It must be noted, however, that judgments of whether something is *really* some kind of thing are themselves not independent of language. Asking the question of whether something *really* is an *X* inevitably requires identifying the potential kind *X* by means of a word. As such, the question may activate thoughts of properties closely associated with the word and the suitability of use of the word for the current context, and judgments may reflect the fit of the object in question to those properties or conditions. In that case, judgments of what something *really* is may not provide evidence of perceiver-independent groupings. Instead, they may reflect the pragmatics of word use.

Malt and Sloman (2007) provided some data on the nature of *really* judgments in experiments evaluating Bloom's (1996) proposal that the creator's intended kind membership serves as an essence for artifacts. They used scenarios in which an object was created under one intention and then given a new use, and a character in the story needed to talk about the object to someone else. (For instance, an object was made to be used as a tea kettle and then adopted for use as a watering can; cf. Matan & Carey, 2001.) They showed that participants judged the name associated with the second use as well as the first to be acceptable for talking about the object, with the relative preferences depending on factors such as the length of time of the second use, which use the addressee was more familiar with, and how much the object resembled objects normally called by each name. They then asked other participants to read similar scenarios and to judge whether the object was *really* an instance of each name. They found that judgments of what the object *really* was tended to favor the name associated with the creator's intention more heavily, but not for all types of scenarios. The bias toward the original name for *really* judgments was modulated by the same pragmatic factors that influenced the name acceptability judgments. Malt and Sloman concluded that an object's history, including its creator's intention, is an important element of how people think about and refer to the object, but it does not serve as an essence that confers membership in a kind. The naming and conceptualization of artifacts are both more pragmatic in nature.

Malt and Sloman's (2007) experiments were designed to elicit considerations of the objects within naturalistic contexts. They therefore used scenarios that described specific uses of the objects and conversational interactions about them. These features may have induced sensitivity toward pragmatic factors even when participants were asked what the objects *really* were. The goal of the present two experiments was to better understand what drives judgments of what an artifact *really* is, and what these judgments can reveal about how people think about artifacts. In particular, we used more neutral judgment contexts that did not engage participants in thinking about specific instances of use of the objects or in conversational interactions based on those uses. We asked whether, under such conditions, *really* judgments would reveal a language-independent kind membership linked to a creator's intention, as the essentialist account

would predict, or whether they would still be modulated by other, more pragmatic influences.

Two interpretations of the second possible outcome can be made. First, such kinds do not exist, and what an artifact *really* is boils down to nothing more than how well its properties correspond to those associated with certain words. Second, such kinds may exist, but judgments of *really* are inadequate to determine whether this is true, because these judgments do entail language and are still contaminated by the conditions of word use. In the General Discussion, we will take up the implications of each interpretation, as well as whether such an outcome would be compatible with a more general notion of artifact kinds.

Experiment 1

In the first experiment, we manipulated how typical the objects' physical features were of things normally called by the name that we queried for each. We also manipulated whether participants believed that the name was given to the object by its creator (with implications for both its intended kind membership and intended function; we did not attempt to disentangle these) or by someone who had no direct knowledge of the creator's intention. For the case in which the creator's intention was known, information about the object, including its name, was provided from a website described as that of the manufacturer of the object.[1] For the case in which the creator's intention was not known, the same name was represented as being bestowed by someone who found the object at a garage sale and posted it for resale on eBay. If the essence-based-kinds view is correct, when the original intention is known, this information should govern judgments of what the object really is. According to this view, the creator's intention endows the object with the essence of a kind, and people act as essentialists when thinking about artifacts. Provided that the intention is known, the object should be judged to be *really* a member of the kind. Only when the original intention is unknown might typicality have a substantial impact. In that case, object features would be the best indicator of the creator's intention, and features more closely associated with the name might be perceived as providing better evidence of the intention. In contrast, if the pragmatics perspective is correct, higher-typicality items might be judged as *really* being an *X* more than lower-typicality ones, regardless of knowledge about the original intention. According to this view,

questions about whether something is *really* an *X* activate thoughts of properties most closely associated with the queried word, and the judgments may reflect the fit of the object in question to those properties. If so, judgments of what something *really* is may not provide evidence of perceiver-independent groupings. Instead, they may reflect the pragmatics of word use. Knowledge of the intention may have some impact on the judgments, because the history of an object matters in people's thinking about the object and useful names for it (e.g., Gelman & Bloom, 2000; Gutheil, Bloom, Valderrama, & Freedman, 2004; Malt & Sloman, 2007), but such knowledge will not fully determine them.

In addition to the *really* judgments, a separate group of participants judged whether each object was *not really* an *X*, where *X* was the same name as in the *really* judgment. This task was aimed at further testing what judgments of *really* tap. If they reflect beliefs about the presence or absence of an essence, objects receiving positive judgments from the first group, as *really* being an *X*, should receive negative judgments from the second group, of *not really* being an *X*. If the ratings reflect how well objects embody the features normally associated with a name, objects of low to moderate typicality may be judged both as *really* being an *X* by the first group and as *not really* being an *X* by the second, because the objects only partially embody the features normally associated with the name. (The higher-typicality objects do not discriminate between these views, because both would predict that such objects would be judged as being *really* an *X* but not *not really* an *X*.)

We also created a manipulation check, by asking additional participants to judge the suitability of the offered name for speaking to the manufacturer's Customer Service representative about the product. If participants noted the manipulation of name source, they should rate low-typicality names as being more suitable for this use when they were provided by the manufacturer than when they were provided by an eBay user.

Part 1a: Typicality judgments

To select stimuli and provide typicality ratings to compare to *really* and *not really* judgments, we first gathered typicality ratings for a set of 28 artifact pictures. From this set, 20 objects were chosen for use in Part 1b.

Method

Participants A group of 16 Lehigh University undergraduates participated.

Materials and procedure The 28 object pictures were collected from company websites offering objects for sale under one of the following names: *can, jar, box, stool, bench, ladder,*

---

[1] Manufacturer names might not always align with the intentions of objects' creators; this type of case may be most likely when an inventor or artisan independently creates an object and then interests a larger entity in selling it, perhaps turning over control of how it is advertised. For the commercial products used here, though, the distinction between a creator and the company making and selling the object was minimal. The objects were of the sort developed by people working directly for the company in response to its need for specific products.

pail, *bucket*, *tape*, or *chest* (of drawers). Two or three examples of each name were chosen, such that the exemplars were likely to vary in perceived typicality while still constituting literal applications of the name. Figure 1 illustrates two instances of *stool*. Other stimuli included a wooden bedroom chest versus a lightweight plastic storage chest, a tall ladder versus a short three-step ladder, canning jars versus a decorative jar with a spigot, and so on. We avoided objects named by conventional compounds such as *juice box* or *hairbrush*, which might be argued to belong to kinds distinct from those named by the head noun alone.

Participants were given written instructions explaining that words can be used in many ways besides their most ordinary use. The instructions gave examples of nonhuman cases of *running*, such as a spider running and water running, as well as of *gun* being applied to a toy gun and a glue gun. These examples were chosen to provide compelling illustration of the diversity of how words can be used and to motivate the interest in judgments of typicality (and *really*, in Part 1b). The actual stimuli to be judged, of both high and low typicality, were all concrete objects to which the names applied in a literal sense, as is illustrated in Fig. 1 and described above. The instructions went on to say that, given all this variety in how words are used, we were interested in intuitions about how typical some objects are of particular names. The instructions then indicated that the participant would see 28 pictures of ordinary objects along with the names that their manufacturers had given them, and that for each one the participants should consider how typical each object shown was of things that get called by that name. Participants were asked to make their responses on a 1–7 scale, with 7 being for *objects very typical of things called by the specified name*, and 1 being for *objects not typical of the name*. Numbers between these anchors indicated intermediate judgments.

Results and discussion

Mean typicality ratings were computed for the 28 objects. Twenty of the objects were selected in pairs, with one being higher in typicality than the other for each of the ten words

Fig. 1 Low- and high-typicality examples of *stool* used in Experiment 1

listed in the Method section above. The mean rating for the higher-typicality items was 6.36 ($SD = 0.44$, range 5.69 to 6.94), and that for the lower-typicality items was 3.50 ($SD = 0.95$, range 1.88 to 4.63).

Part 1b: *Really* and *not really* judgments

Method

Participants A group of 46 Lehigh University undergraduates made *really* judgments: 24 for *really* judgments with manufacturer information present, and 22 for the eBay context. Another group of 48 made *not really* judgments, 24 in each context.

Materials and procedure The 20 objects selected in Part 1a were used. Ten fillers were added, and these were chosen to provide cases that would allow participants to use the middle to lower end of the *really* judgment scale, so that they would not artificially lower ratings on any target objects just to use more of the scale. The fillers included objects sharing only a single salient feature with things normally called by the name (e.g., a robotic vacuum cleaner called *floor maid* by the manufacturer) and objects normally called by a better-fitting or more informative name (e.g., a toy chest that we dubbed a *toy keeper* by editing the webpage information).

The participants who were to make *really* judgments were given written instructions explaining that words can be used in many ways besides their most ordinary way. Similar to the typicality instructions, these instructions gave examples of *running* such as a spider running and water running, and of *gun* as applied to a toy gun and a glue gun. The instructions went on to say that it might be debatable whether these were *really* cases of *running* or *gun*, that some people might think yes and others no, and that some might feel that some cases were and others were not.

The instructions for participants receiving stimulus packets with the creator's intention present then said that the objects would be shown with information about what the manufacturer called them. The instructions for those who would not know the creator's intention indicated that the objects had been found at garage sales by someone who was reselling them on eBay, using the best name that he or she could come up with.

Both versions then said that for each object, the participant should consider how sensible it was to say that the object was *really* an instance of the name that it had been given and to make a response on a 1–7 scale, where 7 meant *it is very sensible to say that the object is really an instance of the name* and 1 meant *it is not very sensible to say that*, and the numbers in between indicated intermediate judgments. The instructions emphasized that it was fine if the participants' judgments fell mostly at one end or the other,

or in the middle, or were a mix from high to low, and that they should just call each stimulus as they saw it.

Each participant then received a booklet that instantiated one of the two contexts. The booklet with the creator's (manufacturer's) intention conveyed this intention by presenting enough of each object's original webpage to show the full object label (e.g., *heavy-duty steel rolling ladder*) and purchasing information (e.g., company name, item price, item details), along with the object. The booklet version lacking intention information presented the same object images stripped of all webpage content, with the same full label typed (by us, in a font that was uniform for all pictures) beneath it. Responses were made on a separate sheet with *It's really an X* followed by a blank for the rating for each object, where *X* was the head noun (e.g., *ladder*) of the object's name. Pictures were presented in two random orders for each type of booklet.

The participants making *not really* judgments were given instructions, booklets, and response sheets that were the same, except for alterations throughout to indicate that the issue was about whether some things were *not really* an *X*.

Results and discussion

Table 1 presents the mean ratings for the higher- versus lower-typicality objects for each context and rating type.

If *really* judgments reflect membership in essence-based kinds, typicality might influence judgments made in the eBay context, but they should not influence judgments in the manufacturer context. Table 1 shows that, contrary to this possibility, the ratings differed for higher- versus lower-typicality objects but were virtually identical for the manufacturer versus eBay contexts. An analysis of variance (ANOVA) with Context as a between-subjects factor and Typicality as a within-subjects factor confirmed a large main effect of typicality, $F(1, 44) = 386.7$, $p < .001$, but no effect of context, $F(1, 44) < 1$, and no interaction between the two factors, $F(1, 44) < 1$. The predicted typicality effect held for all item pairs. This outcome indicates that *really* judgments

of this sort are not substantially influenced by knowing the creator's intention. Instead, they are strongly influenced by the typicality of the object with respect to the name being queried, regardless of what was known about original intentions.

To further explore the impact of typicality on *really* judgments, the mean *really* rating for each object was correlated with its typicality value from Part 1a. The values showed remarkably close correspondence: $r = .98$, $p < .0001$, for the manufacturer context, and $r = .96$, $p < .0001$, for the eBay context. This outcome supports the idea that participants were responding to the extent to which the objects had features matching those brought to mind by the queried word as they made *really* judgments.

The *not really* judgments similarly showed a lack of impact of knowing the creator's intention. An ANOVA confirmed a main effect of typicality, $F(1, 46) = 169.89$, $p < .001$, but not one of context, $F(1, 46) = 1.01$, $p > .3$, and no interaction, $F(1, 46) = 3.14$, $p = .083$.

The main purpose of the *not really* judgments was to evaluate whether some objects would be judged as being both *really* and *not really* an *X*, a pattern that would be incompatible with the idea of *really* judgments revealing membership in essence-based kinds. For *really* judgments, the lower-typicality objects were rated just below the midpoint of the scale for both contexts. For *not really* judgments, the lower-typicality ones were similarly rated close to the mid-point (just below the midpoint for the manufacturer context, and just above for the eBay context). In both contexts, the *really* and *not really* judgments summed to more than 7 (the top of the rating scale), showing superadditivity. These ratings indicate that participants were moderately sure that these objects were *really* examples of the queried name, and also moderately sure that the objects were *not really*.[2] The typicality ratings from Part 1a correlated with *not really* judgments at $r = -.95$ for the manufacturer and $r = -.96$ for the eBay context, $ps < .001$, indicating that *not really* judgments were also strongly influenced by how much the object resembled the things most closely associated with the name.

In sum, in the *really* judgment task used here, which had no scenarios describing interactions with objects or conversational partners, the typicality of the objects with respect to the names queried still had a large impact on the judgments. Knowledge about the creator's intention had little impact. Objects of low to moderate typicality with respect to a name

Table 1 *Really* and *not really* mean ratings (and standard deviations) for higher- and lower-typicality objects in the manufacturer and eBay contexts (Exp. 1, Part 1b)

	Typicality	
	Higher	Lower
Really		
Manufacturer	6.45 (0.45)	3.90 (0.84)
eBay	6.39 (0.39)	3.87 (0.81)
Not Really		
Manufacturer	1.76 (1.18)	3.81 (.99)
eBay	1.65 (.73)	4.35 (.89)

---

[2] Because the midpoint of the scale was not labeled, the participants responding to *really* and *not really* questions might have given it different interpretations. To address this possibility, we collected *really* and *not really* ratings in the manufacturer version from new groups of participants ($n = 22$ for each), with the midpoint labeled "moderately sensible" for both. The results were virtually identical to those reported here.

were considered both *really* an example of the name and *not really* an example. These results favor the pragmatics view of the nature of judgments about what an artifact *really* is.

Part 1c: Creator's intention manipulation check

Method

Participants A group of 32 Lehigh University undergraduates made judgments of the sensibility of the names for calling customer service, with 16 participants in each context (manufacturer versus eBay).

Materials and procedure The instructions from Part 1b were altered to indicate that the study was about what names individuals would prefer for objects in specific contexts. Examples were given of how choices might differ across individuals and contexts. The information about the sources of the names accompanying the pictures in the packet (manufacturer versus eBay) was the same as in Part 1b. Participants were then asked to imagine that they had acquired each specified object as indicated and needed to call the manufacturer's Customer Service department to ask a question about the object. Participants were told to judge how likely they would be to use the name specified. The stimulus packets were the same as in Part 1b. Responses were made on a 1–7 scale, where 7 indicated *high likelihood* and 1 indicated *low likelihood*. The response sheet used the sentence frame "I'm calling about my *X*," where *X* was the same head noun used in Part 1b plus one or two of its accompanying modifiers from the stimulus packet (e.g., "I'm calling about my hexagon glass jars"), for general pragmatic appropriateness. (Pilot work had shown that no one thought that it was suitable to call Customer Service and merely say "I'm calling about my jars.") After completing all ratings, the participants were asked to write down who had given the original name with the pictures.

Results and discussion

In total, 75 % of the participants in the manufacturer context and 88 % of the participants in the eBay context correctly identified the source of the name accompanying each picture. Table 2 presents the mean ratings for the higher- versus

Table 2 Customer service name sensibility mean ratings (and standard deviations) for higher- and lower-typicality objects in the manufacturer and eBay contexts (Exp. 1, Part 1c)

| | Typicality | |
	Higher	Lower
Manufacturer	4.69 (0.94)	4.63 (0.75)
eBay	5.28 (0.82)	4.46 (0.78)

lower-typicality objects for each context and rating type. The ratings for the eBay context showed an impact of typicality, but the ratings for the manufacturer context did not, consistent with recognition by participants of the name source and of its pragmatic relevance to their choice of a name for communicative purposes. An ANOVA confirmed a main effect of typicality, $F(1, 30) = 7.5$, $p < .01$, no main effect of context, $F(1, 30) < 1$, and a significant interaction, $F(1, 30) = 5.37$, $p = .03$. The sensibility ratings did not correlate significantly with typicality in the manufacturer context ($r = .18$), but they did in the eBay context ($r = .63$, $p < .005$). Thus, the results of Part 1b cannot be attributed to a context manipulation that was insufficient to impact judgments within this sort of paradigm.

Experiment 2

In the second experiment, we also presented objects in the absence of scenarios evoking specific physical or conversational interactions. Again we varied how closely the objects to be judged were associated with certain names, but we added a new dimension to this relationship by varying how recently the particular object has been introduced into American culture, and how entrenched that particular object–name relationship consequently would be to younger versus older people.

The set of objects called by a given artifact name evolves over time. For instance, modern telephones, whether land lines or cell, have limited resemblance to telephones of the late 1800s. But at any given moment in history, some traditional and newer versions of things called by the same name may coexist. In this experiment, we asked for *really* judgments of artifacts that varied in how recently they had come into existence (e.g., a dial telephone versus a cell phone). We asked college students and older adults (above age 70) to look at pictures of the objects and to judge whether each was *really* an *X*. If each object is endowed with a kind membership by its creator, then that object should be judged as being *really* a member of the kind to the same degree, independent of its entry point in the culture and, in particular, of whether the person judging happens to have experienced more traditional or newer examples of the kind throughout their life. On the other hand, if *really* judgments reflect the extent to which an object is compatible with the properties most readily brought to mind by a queried name, then people of different age groups may make different judgments. In particular, while being aware of the same range of objects, the age groups may differ in how much they associate newer instances with the name. For instance, *phone* might bring cell phones to mind much more for younger than for older people. Age might then interact with the recency of objects in *really* judgments: Newer versions

may seem to be more *really* examples of the name to younger than to older participants, while judgments may differ less for longstanding, traditional instances.

We also collected typicality judgments from a separate set of college students to evaluate whether perceived typicality for this group would predict *really* judgments for the other student group. Due to the difficulty of obtaining older participants, we did not collect typicality judgments from older adults. However, it will be of interest to see whether college student judgments predict older-adult *really* judgments as well as they do college students'. If older adults' *really* judgments differ from college students' *really* judgments because of somewhat different associations of object properties with names, then college student typicality judgments should correspond less well to the older-adult *really* judgments.

Part 2a: *Really* judgments

Method

Participants A group of 16 Lehigh University undergraduates participated for course credit. In addition, 33 older adults were recruited by undergraduate research assistants and participated without compensation. Most were relatives of the research assistants or friends of the relatives, and they were tested at home during school breaks. A smaller number were employees at the workplace of a research assistant or were approached at intermissions at musical performances at Lehigh. Detailed demographic information was not collected, but the older adults were generally middle-class and college-educated. All were believed to be cognitively intact. To avoid contaminating judgments by making them aware that their age was of interest, ages were not determined before participation was solicited. After completing the judgment task, participants were asked to turn over their sheets and to circle an age range from among the set 51–60, 61–70, 71–80, and 81–90. About half the total solicited fell into the 51–60 or 61–70 age ranges. Because our interest was in people whose experience with objects would be most different from the college students', the participants included for analysis consisted of the 16 adults whose ages fell in the 71–80 and 81–90 ranges.

Materials and procedure The participants were given written instructions similar to those in Experiment 1, except that no mention was made of manufacturers or of selling the objects on eBay. Instead, the instructions simply said that participants would see pictures of ordinary objects, along with a name that each is often given. The 1–7 response scale was the same as in Experiment 1.

The 22 target pictures were collected from websites to form 11 pairs, such that one member of each pair was an object in more common use several decades ago, and the other was called by the same name but had more recently been introduced. The pairs (in the order old, new) were *books* (hardback, CD), *mailboxes* (metal rural delivery, computer e-mail inbox), *keys* (metal for turning in lock, electronic swipe card), *skins* (shed by a snake, for protecting iPods), *letters* (handwritten on paper, electronic document), *cameras* (SLR, on a cell phone), *slides* (film type in cardboard holder, PowerPoint), *folders* (manila, on a computer), *phones* (land line with dial, cellular smartphone), *rulers* (12-in. wooden, digital), and *pointers* (telescoping metal, laser). Figure 2 shows the two *key* objects. The stimuli were arranged in booklets, mixed among 14 filler items to yield 36 items in total, 12 per page. The 14 fillers were similar in nature to those in Experiment 1, chosen to allow the participants to use the middle to lower end of the *really* judgment scale, without artificially lowering ratings on the target objects so as to use more of the scale. Two random orders of items were used. To introduce the target name to be judged, above each picture was a statement indicating that a common user group of the object *calls the object an X* (e.g., *Hotel patrons call this a key*; *Librarians call these books*; *Computer users call this a mailbox*).

The response sheets were similar to those in Experiment 1. We created two versions of the response sheet corresponding to the two random orders of the pictures. After the main task was completed, the older adults were asked to turn over their response sheets and circle an age range, as described above, and their native language. For the last 11 participants (who included eight of those 71–90 years of age whose data were used for the analysis), we added four questions that asked about use of the Internet, e-mail, cell phones, and digital cameras. The participants circled *yes* or *no* to indicate whether they used each one regularly, with *regularly* defined as at least a few times a week.

Results and discussion

Mean *really* ratings were computed for old and new items and for each age group. As Table 3 shows, the older-adult judgments were slightly lower than college students' for old items, but were more substantially lower for new items. An ANOVA with Age Group as a between-subjects factor and Object Type as a within-subjects factor showed a main effect of age group, $F(1, 30) = 5.16$, $p < .05$, and a main effect of object type, $F(1, 30) = 198.37$, $p < .001$. Importantly,

Fig. 2 Old and new examples of *key* used in Experiment 2

Table 3 *Really* mean ratings (and standard deviations) for old and new objects among college students and older adults (Exp. 2)

Participant Age Group	Object Type	
	Old	New
College student	6.25 (0.46)	4.50 (0.96)
Older adult	6.13 (0.53)	3.64 (0.88)

we also found a significant interaction of object type with age, $F(1, 30) = 5.96$, $p < .05$. Consistent with the idea that *really* judgments reflect the extent to which an object is compatible with the properties most readily brought to mind by a queried name, older adults' judgments differed from those of younger adults more strongly for just those objects for which the two groups' experience was likely to be most different.

Of the eight older adults who were asked to report their use of the Internet, e-mail, cell phones, and digital cameras, one failed to respond. Among the remaining seven, all said *no* to regular use of a digital camera, but five said *yes* to regular use of a cell phone, and three each said *yes* to regular use of the Internet and e-mail. If we can extrapolate this subset to the rest of the older-adult group, it seems that even with a sample that includes participants who are active users of some of the queried types of objects, the perception by older adults of what objects are *really* examples of *X* differs from that of the younger generation.

The middle of the rating scale, 4, was labeled "moderately sensible" (to say that the object was *really* an *X*), and the older-adult mean across the new items was close to this value at 3.64. So, it does not appear that this group was simply rejecting the new items out of hand due to unfamiliarity with them or the names used for them. The larger difference between the old and new item ratings for the older adults held for seven of the 11 items tested. Most noteworthy are two cases in which the ratings actually reversed between college students and the older adults. Older adults rated carousel-type slides slightly above the midpoint of the scale (4.75) but rated PowerPoint slides slightly below it (3.56). College students also rated the carousel-type slide slightly above the midpoint (4.88), but they rated PowerPoint slides higher (5.44). Likewise, older adults rated the dial phone close to the top of the scale (6.75) and the cell phone near the midpoint (4.31). College students also rated the dial phone high (6.38), but they rated the cell phone even higher (6.63). For these items, the dominant association of the target name appears to have reversed for college students as compared to the older generation, following the generational shift in the primary applications of the queried names to types of objects.

The full set of adults 50 years old and above from whom data were collected showed a trend toward the same interaction (with a mean rating of 6.08 versus the college mean of 6.25 for old items, and a mean of 3.97 versus the college mean of 4.5 for new items), but the interaction fell short of significance. The subset of participants 71 and above was only half the size of the full sample, but it did show a significant interaction. People in their 50s and 60s, especially those sampled on a university campus, are often working adults who make extensive use of cell phones, computers, and other recent technology on a daily basis. Their associations of objects with names thus may be much more similar to those of college students. The oldest participants would have spent more of their lives interacting with older versions of the objects and less with current versions, and their associations thus would differ the most.

Part 2b: Typicality judgments

Typicality judgments for the same objects were collected in order to assess their relation to *really* judgments.

Method

Participants A group of 22 Lehigh University undergraduates who had not participated in Part 2a participated for course credit.

Materials and procedure The instructions were brief, saying only that pictures of 36 objects would be viewed and that participants should judge how typical each one was of the name given. The responses were made on the same 1–7 scale used for the typicality judgments in Experiment 1. The booklets presenting the pictures were similar to those for the *really* judgments, except that above each picture was only the name to be judged, not a statement that some group of people called the object by that name. The responses were made on a sheet similar to that for the *really* judgments, except that instead of asking whether each object was *really* an *X*, only the name was listed, followed by space for the rating.

Results and discussion

Mean typicality judgments were computed for each of the objects. The ratings correlated with the college student *really* judgments at $r = .91$, $p < .001$, confirming the strong relation of *really* judgments to perceived typicality found in Experiment 1. The mean ratings for the college student typicality judgments correlated with *really* ratings by the older adults at $r = .77$, $p < .001$, a marginally significantly lower correspondence ($z = 1.56$, $p < .06$), indicating that college students' typicality perspectives on objects are less effective at predicting older adults' beliefs about what

objects really are.[3] The difference supports the idea that the divergence in older-adult judgments is mediated by a difference in the perceived typicality of the objects with respect to the names.

In sum, the results of this experiment again favored the pragmatics view of judgments about what an artifact *really* is. Judgments were made in the absence of any specific communicative context, but the typicality of the objects with respect to the name queried still had a large impact on them. Most importantly, judgments of more recently introduced artifacts differed between younger and older participants, indicating that the same object could be judged as being more or less *really* something depending on the entrenchment of the object–name relationship.

General discussion

Experiment 1 showed that people's judgments of what an object *really* is closely reflect their perception of how typical the object is with respect to the queried name. Furthermore, objects of low to moderate typicality were considered both *really* and *not really* the same thing to about the same degrees. Experiment 2 showed that *really* judgments can also vary as a function of age, reflecting, presumably, the fact that different frequencies of mapping names to types of objects result in different perceptions of the typicality of an object with respect to a name. In contrast to the large impact of an object's properties relative to those associated with the queried word, we found no evidence that the creator's intended function or kind membership for the object substantially influenced judgments.

The goal of the experiments was to better understand what drives judgments of what an artifact *really* is, and what these judgments can reveal about how people think about artifacts. The data clearly favor the pragmatics view over the kinds view as an account of what drives *really* judgments. What, then, do the data reveal about how people actually think about artifacts? There are two possibilities, as we identified earlier. First, it may be that people do not treat artifacts as belonging to essence-based kinds. Second, it could be that people do treat artifacts as belonging to essence-based kinds, but judgments of what something

---

[3] One might wonder whether the lesser correlation for older adults was simply due to noisier data. We calculated the standard deviations of the ratings for each item among college students and older adults. The mean standard deviations across all items were 1.23 for college students and 1.43 for older adults. This small difference was due mainly to the new items (with a difference in mean standard deviations of 0.31, as compared to 0.15 for the old items), consistent with the basic phenomenon itself. Older adults were highly similar to the college students in variability on the more entrenched object–name pairings, but the groups differed a bit more for the recent ones, for which their individual levels of experience were likely to vary more.

really is cannot reveal this, because the judgment taps linguistic rather than pure kind knowledge.

Along the lines of the second interpretation, Armstrong, Gleitman, and Gleitman (1983) found that people made graded typicality judgments for entities that should be clear-cut examples of certain terms. For instance, the number 3 was judged as being a better exemplar of *odd number* than was 57, even though both equally well meet the mathematical criterion for being an odd number. Armstrong et al. attributed the graded responses to the familiarity/frequency of the entities and of their features as exemplars of the queried names. Possibly the present results represent a similar phenomenon for *really* judgments, with nongraded beliefs about kind membership being obscured in the judgments by more superficial aspects of experience. For Armstrong et al.'s research, the interpretation was straightforward because their stimuli (*odd number*, *even number*, *female*, and *plane geometry figure*) were selected as having necessary and sufficient conditions for application given by authoritative sources (such as mathematicians or biologists). For artifacts, the problem is that no independent definition is available of what might belong to the relevant kinds, if they exist. The burden is on researchers who want to argue in favor of essence-based artifact kinds to find a way to identify kinds and to verify their status as essence-based. It seems that asking questions involving names for the objects will not be among tasks that might be helpful in that enterprise, even when the judgments are made outside of a conversational context.

The present data may be compatible with the existence of artifact kinds in the more general sense—namely, clusters based on multiple shared properties. If those clusters are assumed to have fuzzy boundaries, and if *really* judgments, like generic categorization judgments (e.g., *Is it a chair*?; see, e.g., Rosch et al., 1976) are simply judgments of how closely the object matches some center or prototype of the named cluster (e.g., Hampton, 1993), then the present data do not directly speak against this sort of notion of artifact kinds. There remains, however, the problem of cross-linguistic variability in naming. Given that languages have different ways of grouping artifacts by name, and that the named groupings do not have simple superset–subset relations to one another, it seems that the properties of artifacts do not always create clear clusters of objects that are "intrinsically separate" (Rosch et al., 1976). This possibility is supported by the scaling solutions on similarity data for containers in Malt et al. (1999), in which the objects spread out across conceptual space, with some objects clustering to varying degrees and others falling into the spaces between these clusters. Accommodating this situation in terms of kinds would require that some familiar, ordinary artifacts either belong to no kind (on the basis of their position in conceptual space) or else change their kind, depending on

the language spoken. Both options seem at odds with the idea of objects having membership in kinds.

What may make more sense is to describe the artifact case simply as objects spreading out across conceptual space with varying degrees of clustering, and to leave the account at that. This approach is not greatly different from the general sense of kinds, except on one critical point: It does not entail the existence of any discrete, nonlinguistically defined kinds. It still allows for the very real fact that people have intuitions about objects belonging to kinds, and that people frequently ask questions such as *What is that*? or *What kind of thing is that*? However, it allows that the answers to the questions will be language-dependent. An object for which the answer is *bottle* in English may be a *mamadera* rather than a *botella* in Spanish, and so on. Our account also accommodates the observation that an object may be judged to be *really* or *not really* an *X* simply to the extent that it has the properties most associated with the word *X* in the language of test.

This proposal helps solve some puzzling questions that arise if one tries to pursue the argument that artifacts do *really* belong to some kind that may differ from what their name suggests. For instance, if an electronic mailbox isn't *really* a mailbox, and a magnetic swipe key isn't *really* a key, then what are they really? And why do we call them *mailbox* or *key* if they are *really* examples of something else? One could say that they are only metaphorically mailboxes and keys, but this response does not fully answer either of those two questions. At the same time, electronic mailboxes and magnetic swipe keys seem less metaphorical than cases such as tripod feet and bed skirts. Where would the dividing line fall between things that are *really X* and things that are only metaphorically *X*? Likewise, one could argue that in Experiment 1, there might have been a creator (designer) of the objects who intended some of them to belong to a different kind than that reflected in the name given on the manufacturer webpage. If so, what are they really, and why were they named after something that they really were not? Choices among names can be made strategically, for sales purposes. But that observation only implies that several possible names might be acceptable given the object features, and that one of the names has more useful or desirable associations for sales purposes. It does not suggest that the objects are not what they are called. To pursue the argument that objects are *really* something else, one would have to identify kinds that the creator *really* meant the objects to be, distinct from the kinds that they were named for. Recognizing instead that things can be perceived as *really* being an *X* to varying degrees, depending on objects' relations to the properties associated with *X*, accounts for the intuitions about whether they really are or are not what they are called, without needing to identify some "true" kinds that they belong to.

This proposal also helps explain other, related intuitions. Some people report that they feel that the newer objects in Experiment 2 might belong to different kinds than the older ones, or that certain pairs of old and new instances might share an essence while others do not. For these intuitions to reflect reality, there would have to be identifiable other kinds that some of the new instances belong to that are not reflected in their names, and that have somehow come into being despite the intention of a creator to make a variant of an older object that serves much the same purpose and can be called by the same name. It seems more likely that such intuitions reflect the extent to which the old and new examples of a name overlap with typical instances of the name or with each other. For instance, the intuition that traditional books and electronic books may belong to different kinds seems to reflect primarily their great physical differences and the secondary usage details that follow, since the most central element of their intended use (to provide for the reading of lengthy connected discourse), the content that they convey, and their names remain the same.

This conclusion is broadly compatible with other research suggesting that people treat artifacts differently from natural kinds and do not treat artifacts as if they have essences (e.g., Gelman, 1988; Hampton, Storms, Simmons, & Heussen, 2009; Kalish, 1995, 2002; Malt, 1990; Rhodes & Gelman, 2009). However, it goes beyond such literature, to question the notion of artifact kinds per se, and to suggest that intuitions about kind membership are in reality intuitions about the relation of objects to properties associated with the words of a language.

Conclusion

It is widely assumed in the psychological and philosophical literature that artifacts fall into distinct kinds. These kinds are generally identified by appeal to words—*chair*, *stool*, *bowl*, *vase*, and so on—which raises problems for the notion of artifact kinds once contextual and cross-linguistic variation in the sets of artifacts grouped together by name are recognized. Possibly, judgments of what artifacts *really* are would reveal their true kind membership, as distinct from what they are called in communicative contexts. However, we found that people failed to treat artifacts as having a definitive kind membership in their judgments of what objects *really* are. Instead, *really* judgments reflected the typicality of the objects with respect to things normally called by the queried name. If these judgments are taken as direct evidence about the existence of artifact kinds, the outcome argues against there being such kinds. An alternative interpretation is that *really* judgments are fundamentally linguistic in nature, and so do not tap into underlying kind memberships. In either case, if such kinds exist, they remain to be found, using tasks independent of linguistic influence

on judgments. A more likely reality may be that intuitions about the existence of artifact kinds reflect the partial clustering of objects in similarity space and the fact that each language provides names for constellations of objects in that space.

References

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53,* 60–80.

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

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60,* 1–29.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bloom, P. (2007). More than words: A reply to Malt and Sloman. *Cognition, 105,* 649–655.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science, 20,* 252–257.

Elder, C. L. (2007). On the place of artifacts in ontology. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 33–51). Oxford: Oxford University Press.

Futo, J., Téglás, E., Cisbra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition, 117,* 1–8.

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). Children are sensitive to how an object was created when deciding what to name it. *Cognition, 76,* 91–103.

German, T. P., & Johnson, S. C. (2002). Function and the origins of the design stance. *Journal of Cognition and Development, 3,* 279–300.

Grandy, R. E. (2007). Artifacts: Parts and principles. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 18–32). Oxford, U.K.: Oxford University Press.

Gutheil, G., Bloom, P., Valderrama, N., & Freedman, R. (2004). The role of historical intuitions in children's and adult's naming of artifacts. *Cognition, 91,* 23–42.

Hampton, J. A. (1993). Prototype models of concept representation. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive analysis* (pp. 67–95). London, U.K.: Academic Press.

Hampton, J. A., Storms, G., Simmons, C. L., & Heussen, D. (2009). Feature integration in natural language concepts. *Memory & Cognition, 37,* 1150–1163. doi:10.3758/MC.37.7.1150

Hauser, M. D. (1997). Artifactual kinds and functional design features: What a primate understands without language. *Cognition, 64,* 285–308.

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

Kalish, C. W. (2002). Essentialist to some degree: Beliefs about the structure of natural kind categories. *Memory & Cognition, 30,* 340–352.

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

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

Kemler Nelson, D. G., Herron, L., & Morris, M. (2002). How children and adults name broken objects: Inferences and reasoning about design intentions in the categorization of artifacts. *Journal of Cognition and Development, 3,* 301–332.

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extentionist semantic view. In J. W. D. Dougherty (Ed.), *Directions in cognitive anthropology* (pp. 91–113). Urbana, IL: University of Illinois Press.

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 340–373). Washington, DC: Georgetown University Press.

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29,* 289–315.

Malt, B. C., & Paquet, M. (2008). The real deal: What judgments of *really* reveal about how people think about artifacts. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 247–252). Austin, TX: Cognitive Science Society.

Malt, B. C., & Sloman, S. A. (2007). Category essence or essentially pragmatic? Creator's intention in naming and what's really what. *Cognition, 105,* 615–648. doi:10.1016/j.cognition.2006.10.001

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40,* 230–262.

Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition, 78,* 1–26.

Phillips, W., & Santos, L. R. (2007). Evidence for kind representations in the absence of language: Experiments with rhesus monkeys (*Macaca mulatta*). *Cognition, 102,* B455–B463.

Putnam, H. (1975). The meaning of "meaning". In H. Putnam (Ed.), *Mind, language, and reality: Philosophical papers* (Vol. 2, pp. 215–271). Cambridge, U.K.: Cambridge University Press.

Rhodes, M., & Gelman, S. A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review, 16,* 920–924. doi:10.3758/PBR.16.5.920

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. doi:10.1016/0010-0285(76)90013-X

Taylor, M., & Flavell, J. H. (1984). Seeing and believing: Children's understanding of the distinction between appearance and reality. *Child Development, 55,* 1710–1720.

Thomasson, A. L. (2007). Artifacts and human concepts. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 52–73). Oxford, U.K.: Oxford University Press.