INDIVIDUALS AND THEIR CONCEPTS

Sergey Blok, George Newman, and Lance J. Rips

Northwestern University

To appear in: W-k. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.),

Categorization inside and outside the lab. Washington, D.C.: American Psychological Association.

INDIVIDUALS AND THEIR CONCEPTS

Nearly all research on concepts in cognitive psychology is research on categories of objects — categories of teapots or turnips, for example. But when it comes to things that are important to us — people, pets, works of art, special places — we also represent the individuals themselves, not just the categories they belong to. Proper nouns, such as *Herman Melville, Toto, Broadway Boogie-Woogie*, or *Hudson Bay*, can denote these individuals, but you can also represent individuals for whom you have no conventional names, like the bed you usually sleep in or your neighbor's mulberry tree.

It is Doug Medin who is mainly responsible for calling category researchers' attention to the importance of individual concepts. Medin and Schaffer's (1978) Context Model proposed that much of what we know about categories we determine from our memories of their exemplars. It's hard to overestimate the importance of this model: Not only has it produced generations of similar theories of categorization (e.g., Kruschke, 1992; Nosofsky, 1986), but it has also influenced fields as diverse as the psychology of attention (Logan, 2002), social psychology (Smith & Zarate, 1992), and phonology (Pierrehumbert, 2001), to name just a few. In other work, Doug has stressed the more abstract information that concepts afford (e.g., Medin, 1989; Medin & Ortony, 1989), but he still retains a fondness for exemplars. In fact, this tension in Doug's thinking about concepts is characteristic of a special turn of mind, a form of reasoning that we're tempted to call "modus medins" and that the following schema approximates:

P.

Not-P.

I still think there's something right about P.

Not only is modus medins completely valid in classical logic, its conclusion also anticipates our own conclusion about exemplars. Although we may not be able to reduce the representation of categories to

the representation of exemplars, we do, nevertheless, have representations of exemplars whose properties are worth exploring in their own right.

Certainly, concepts of individuals and concepts of categories must interconnect in our thinking. Your neighbor's mulberry tree is a member of the mulberry tree category, a member of the tree category, and so on; we typically represent individuals as category instances. There may, however, be less obvious dependencies than this membership or exemplar relationship. According to certain ancient and modern theories, the concept of an individual depends so tightly on the concept of its categories that the individual's very persistence, identity, and distinctness derive from the category.[1] The concept of a tree, for example, may dictate when your neighbor's mulberry tree begins its life and finishes it, where the mulberry tree's spatial extent ends and the ground begins, and whether the mulberry-as-a-sapling is the numerically same item as the mulberry-as-a-mature-tree. The aim of the present paper is to examine this doctrine, which we will call *sortalism,* and its implications for the psychology of concepts.

Sortalism

A sortal is a general term that, roughly speaking, denotes a kind of thing or category. Many common count nouns, especially basic-level nouns, such as *tree*, *person*, or *book*, are sortals in good standing (see Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, for the classic account of basic-level concepts). According to the doctrine that we are examining here, the meaning of a sortal differs from the meaning of other natural-language predicates in including certain principles that apply to objects falling under them. Sortalists believe that both general terms like *tree* and predicates like *blue* or *wooden* have meanings that specify which objects qualify as instances (a *principle of application*). The meanings of both *tree* and *wooden*, for example, determine that the neighbors' mulberry tree is an instance of both. However, sortals (but not other predicates) also determine the identity of individual instances across times and situations (a *principle of identity*). Knowing the meaning of *tree*, for example, entails knowing in theory not only that certain things are trees, but also that a particular tree at time t_1 is the same as one at

time t_2 (e.g., Dummett, 1981; Geach, 1962; Gupta, 1980; Wiggins, 1980). *Wooden*, however, yields no such information about its instances.

Some sortalists also separate the principle of identity from a *principle of individuation* whose job is to distinguish instances for purposes of reference or enumeration. For example, the principle of individuation for *tree* determines whether a group of trees contains, one, two, or more (e.g., Macnamara, 1986).[2] Sortals are important because they support crucial aspects of our knowledge of individuals. The ability to count objects, for example, is supposed to depend on the presence of a sortal to individuate the objects in a definite way. "To see the logical role sortals play in our thought, first consider that we cannot simply count what is in this room. Before we begin counting, we must be told what to count. We must supply a sortal..." (Carey, 1995, p. 108). In a similar way, we can't count things that the predicate *wooden* applies to, since *wooden* doesn't supply a principle of individuation. In a given amount of wooden stuff — say, the stuff comprising a wooden table — do we have one wooden thing (the table), five (the legs and the top), six (the legs, the top, and the table), or more? Sortals such as *table* or *leg* seem to resolve this ambiguity and allow counting to take place.

Sortalists also hold that the meaning of proper names presupposes a sortal to which the denoted individual belongs. According to Macnamara (1986), for example, a child who correctly learns the name *Cat-astrophe* for the family cat must know a basic-level sortal (e.g., *cat* or some synonym in the language of thought) that denotes a category for Cat-astrophe. The sortal's principle of identity guides the child's recognition that the name applies to the same individual over the animal's incidental changes in position, perceived shape (during movement), size (during growth), activities, and other properties. The principle of identity for *cat* supplies the notion of sameness that allows the child to ignore these temporary characteristics and to attend to ones that matter for identifying Cat-astrophe. As Dummett (1981, p. 179) puts it, "Mill wrote as though the world already came to us sliced up into objects, and all we have to learn is which label to tie on to which object. But it is not so: the proper names which we use, and the

corresponding sortal terms, determine principles whereby the slicing up is to be effected, principles which are acquired with the acquisition of the uses of these words."

If the principles of identity and individuation are essential parts of the meaning of words like *cat* or *fridge*, then children who haven't mastered these principles haven't attained a complete understanding of these terms' meanings. Xu and Carey (1996) have taken this idea as a touchstone for infants' knowledge of basic-level concepts. These investigators found no evidence that infants younger than about 10 months notice the difference between a display in which a toy elephant disappears behind a screen and later reappears from it and a second display in which a toy elephant disappears behind a screen and a toy truck emerges from it. Xu and Carey argue on these grounds that the infants have no (correct) principle of identity for the elephant or truck and that they therefore have no concept corresponding to *elephant* or *truck*. The empirical evidence behind this conclusion is a point of current controversy (see Wilcox & Baillargeon, 1998, and Xu, in press, for a review). To our knowledge, however, recent psychological accounts have not questioned the assumption that understanding the principles of identity and individuation is necessary for understanding the meaning of general terms such as *tree* or *toaster*.

We note that there are variations within the sortalist camp about which concepts function as sortals, and we will return to this issue in a later section (see *Sorting Things Out*). For the time being, however, we will focus on the claim, common to all sortalists, that individual identity and persistence depend on the individual's sortal.

Psychosortalism

Our concern in this paper is whether adults' intuitions about individual objects match those of sortalism. Do people think that an individual's identity depends on the sortal category to which it belongs? Do they believe, for example, that whether Cat-astrophe-at-10-years is the same as Cat-astrophe-at-5-years depends on membership in the cat category? Because this issue is about people's beliefs, it is one step removed from the metaphysical questions that inspire sortalism (e.g., What is an

individual object? What determines its temporal and spatial boundaries?). It is possible that everyday beliefs about identity or individuation are wrong and that careful consideration of these issues reveals a highly counterintuitive theory to be the right one (see Parfit, 1984, chap. 11). To make room for this possibility, we can take the point at issue to be the status of *psychosortalism*: the doctrine that people think that identity and difference of individuals crucially depend on sortals.

Psychosortalism bears similarities to psychological essentialism in current theories of concepts and categories (Keil, 1989; Gelman & Hirschfeld, 1999; Medin & Ortony, 1989). Psycho-essentialists claim that people's ordinary beliefs about natural categories include the notion that a central mechanism causes the typical properties of category members and distinguishes one category from another. Although people may not always be able to describe this essence, they nevertheless believe there is one. Psycho-essentialism and psychosortalism have some clear interconnections, since some sortalists maintain that membership in a particular sortal category provides an essence for individual instances (see, e.g., Gupta, 1980, on substance sortals, and Wiggins, 1980). For example, membership in the cat kind may determine when Cat-astrophe's existence begins and ends via the principle of identity for cats, and if so, provides an essence for Cat-astrophe and other members of her species. Other proposals suggest that there may be two essences: one that determines identity and the other that determines category membership (Gutheil & Rosengren, 1996). We note, however, that although psychological essentialists tend to express considerable skepticism about essentialism itself (see Gelman & Hirschfeld, 1999; Medin & Ortony, 1989), parallel doubts about sortalism have not appeared in psychological work on this topic.

Sortalism and psychosortalism are distinct theories with distinct subject matter, but of course they aren't entirely independent. Sortalism is based on systematic consideration of how the identity of objects operates through time and on how identity depends on membership in sortal categories. The results of these thought experiments can diverge from untrained intuitions, but there are limits on such differences. We would be uncomfortable with a sortalist theory whose predictions were at odds with ordinary but clear

intuitions about how objects change across circumstances. Evidence against psychosortalism can provide problem cases that sortalists need to accommodate or to explain away.

Antisortalism

Opposed to sortalism is the view that certain objects maintain their identity and distinctness without support from categories. Suppose, for example, that an individual cat is the product of physical causes that bring the cat into existence and preserve her integrity across her life span. Similar causes will, of course, bring about other cats in accord with biological principles, but a cat's integrity and persistence depend only on the participating causes and not on its membership in the cat category. Following Ayers (1997), we can say that this antisortalist alternative sees an object (at least, a physical object, such as a natural kind or artifact) as a bottom-up product of its causal matrix, whereas sortalists claim that objects are always a top-down product of their categories.[3] Antisortalism thus holds the position that Dummett (1981) attributes to Mill in the passage quoted earlier: Some objects, like loaves of airbread, come pre-sliced by causal factors and are then available for labeling with proper nouns. Antisortalists need not commit themselves to the idea that all objects are free-standing in this way. Certain offices (e.g., the Vice President of the U.S.), roles (e.g., husband), or occupations (e.g., carpenter), if they are objects at all, are objects in virtue of top-down social matters, not bottom-up physical forces. But we can distinguish social categories of this kind from the common artifacts and natural objects that we consider here.

We suspect that most psychologists take a similar bottom-up view of natural kind categories (the categories themselves, not their instances). They believe that species like cats or kumquats are kinds in nature, produced by evolutionary and other causal forces. These species do not depend for their existence on human cognizing or language and could continue as kinds — might even exist more abundantly — if there were no humans at all. There is a fact of the matter as to whether a particular instance belongs to a kind, so that further scientific investigation can potentially resolve uncertain cases. We also suspect that the man-or-woman-in-the-street shares this view of kinds (see Rips, 2001, for a review of relevant research). Although they may not share a scientific evolutionary view of natural kinds, they nevertheless

believe that kinds are a matter of fact rather than a human construct. For example, participants judge that an instance which is described as in between two natural categories (e.g., a fish that "seems to you to be sort of halfway between" an anchovy and a sardine) is "probably one or the other, but I don't know which," in preference to "you can think of it as either one" or "it can't really be either one, then" (Malt, 1990). We suspect, in fact, that realism about natural categories rather than essentialism per se is what most psycho-essentialists are tracking.

This type of hardcore realism with respect to natural categories may make plausible a similar antisortal realism with respect to natural individuals. Just as causal forces can shape a natural category as an independently existing species, so causal forces can shape a natural object as an independently existing individual. (For those who believe species *are* individuals, then this analogy amounts to an identity; see, e.g., Ghiselin, 1974; Hull, 1978.) Physical events bring about Cat-astrophe's birth, segregate her from other background objects, ensure her persistence over time and space, and eventually cause her demise. Although some of these events are obviously similar to those responsible for other cats, it is possible to follow Cat-astrophe's life course by following a complex series of causes and effects, without essential reference to her membership in the cat species. Likewise, the free-standing character of Cat-astrophe gives her the credentials she needs to be counted.

Antisortalism of the type we have just described is not the only form of opposition to sortal theories.[4] But as we mentioned a section ago, our aim is not to decide between sortalism and antisortalism. Instead we hope to make it plausible that everyday beliefs about natural objects might treat these objects as bottom-up and free-standing individuals, rather than individuals bound to specific sortal categories. In the rest of this chapter, we present some evidence that this is so.

Antipsychosortalism

Bloom (2000) suggests three reasons for thinking that people do not behave as though they need sortals to have concepts of individuals. First, people can individuate, track, and talk about objects of dubious or unknown kinds. For example, one can follow the path of an object in the night sky without

being certain as to whether it is a satellite, an airplane, or a UFO. Furthermore, people appear to track individuals that belong to a wide range of kinds by attending to spatiotemporal contiguity and not to any attribute unique to a kind. That is, there is nothing in the concept *chair* or *cup* that can inform identity above knowing that chairs and cups are objects and that objects tend to remain in place or move along a continuous trajectory. Finally, transforming an individual into a different kind of thing can preserve its individual identity. For example, people appear to be comfortable with statements like "this post-office used to be an elementary school" or "this bike trail used to be a rail-road." Here, we provide a more systematic investigation of the way adults reason about identity of individuals undergoing transformations that alter their category membership.

We are interested in whether people will use sortals to guide judgments of identity. In the case of a human, let's assume for now that the relevant sortal is *person* or some synonym. So the issue is whether judgments about an individual's identity depend on his or her membership in the category persons. Experiment 1 looks at judgments of both individual continuity and person continuity to see whether these judgments dovetail across possible transformations. For nonhuman objects, the sortal category may not be so obvious, but many common objects have a default category (probably the basic-level category) that people use in saying what the object is. Sortalists believe that this category is the sortal for the object (Wiggins, 1980, 1997), and we can therefore use category labeling as a way of identifying the sortal. Experiment 2 surveys a variety of natural kinds and artifacts to see whether individual continuity tracks sortal continuity for these items.

Our transformation method is similar to one that Liittschwager (1995) used to examine children's and adults' willingness to attribute the same name to people after a change in state. Liittschwager's participants heard stories and saw line drawings of children who were then said to undergo transformations to a variety of possible end states, from minimal transformations of temporary properties (a clean child to a dirty one) to more extreme category changes (a child to a rabbit or a child to smoke). The results of these experiments showed that both adults and children were less willing to give the original child's name (e.g., Ali) to the final product the greater the "transformational distance" between

them (e.g., they were less likely to say that the smoke was Ali than that the dirty child was). However, there was no clear breakpoint on this continuum where naming fell off steeply—in particular, no elbow where the transformation crossed the basic-category level (e.g., child to adult vs. child to rabbit). Liittschwager concluded that the results "provided little evidence for the position of Macnamara (1986) and other philosophers that identity cannot be maintained across changes that alter basic-level kind [the presumed sortal]. Instead, subjects' responses indicate that people have flexible notions about identity that are based on a combination of different cues" (Liittschwager, 1995, p. v). To obtain more positive evidence on this score, we ask participants for separate judgments of individual continuity (still the same individual?) and category continuity (still a person?) to see whether these judgments diverge (Liittschwager stipulated the outcome categories). For a given transformational distance, we also manipulate the way in which the transformation takes place in order to pin down the factors responsible for maintaining identity.

Experiment 1: Individual Continuity and Person Continuity

According to psychosortalists, whether or not a particular individual — let's call him "Jim" — is the same across possible changes should be a function of whether Jim remains a person during these changes. If Jim drops out of the person category, then Jim himself goes out of existence. For example, if Jim ceases to be a person at his death, then Jim himself ceases to exist at this point. (See Experiment 2 for norming data that confirm that *person* is the sortal.)

Antipsychosortalists can maintain, however, that Jim could undergo certain transformations that alter his status as a person but, at the same time, preserve his status as an individual. As one possibility, people may believe that as long as Jim's mind/brain continues to function as before, he's still the same individual (still Jim) even if he undergoes a radical transformation that would make him a nonperson. For example, we describe in Table 1 two scenarios about an accountant named Jim, who is the victim of a serious auto accident. As a result, Jim's brain is transplanted into a new body: either a robot body

(Type 1 transplant) or a human body that is grown as a "spare part" for just such emergencies (Type 2 transplant). In both cases, Jim's original body is destroyed in the process.

Please insert Table 1 about here.

We asked participants to decide whether the result of the operation was still Jim and also whether he was still a person. The main prediction is: If people use sortals to guide identity judgments, then the decision about whether the resulting creature is still a person should correspond to the decision about whether the creature is Jim. There should be no interaction between question type (Still a person? vs. Still Jim?) and whether Jim's brain comes to lodge in a human or a robot body. In contrast, antipsychosortalists would predict that Jim can maintain his identity even when he switches sortal categories. People may be willing to say that the post-op creature is still Jim, though no longer a person, when he debuts as a robot. If so, there should be an interaction between question type and body type (human vs. robot). We assume that proper names like *Jim* are rigid designators, always referring to the same individual across situations or possible worlds (Kripke, 1972). Participants who state that the transplant recipient is no longer Jim are therefore judging that the recipient is no longer the same individual.

In addition to looking at effects of post-op body, we also investigated other factors that may contribute to judgments of identity continuity. For example, people may have an intuition that the continuity of persons depends on the preservation of their autobiographical memories. Philosophers since Locke (1694/1979) have taken continuity of memory as necessary and often sufficient for identity preservation (Shoemaker, 1970). Our goal here is to see whether people are sensitive to memory continuity in judging identity. More importantly, if people use different kinds of information when judging identity and deciding category membership, then memory continuity may matter more for identity than for category judgments. In the experiment, each scenario has two versions: In one, Jim's memories survive intact, and in the other his memories are lost (as determined by whether the bracketed phrases in Table 1 appeared in the story).

At an extreme, people may believe that continuity of memory is *sufficient* for identity. That is, as long as Jim's memories are intact in the new setting, the resulting creature is still Jim. Developmental studies suggest that children learn at about age seven or eight that brains have an important role to play in preserving identity (Gottfried, Gelman, & Schultz, 1999; Johnson, 1990). For example, second graders and older children tend to believe that if their brain was transplanted to a pig, the transplant recipient would tend to answer to the child's name and not the pig's. However, such a brain transplant inevitably entails the transfer of memories, so existing studies are not able to shed light on whether memories alone are sufficient grounds for identity continuity. In the present study, some participants were told about a brain transplant, while others were told that memories from the original brain were copied onto a computer. This computer was then placed in control of a robot or humanoid body. If memory continuity is sufficient for identity, then Jim should survive in this second condition despite lacking his original brain.

Method. We gave all four versions of the scenarios in Table 1 to a group of participants and asked them to decide whether the resulting creature was Jim and whether it was a person. The characters in the four scenarios had different names (Jim, Ken, Edward, and Bill) in order to distinguish these problems, but the scenarios were otherwise the same as in Table 1. Each scenario appeared on a separate page of a booklet, and following the scenario was a page of questions. For each version, the questions asked participants to rate their agreement or disagreement with a statement about individual continuity ("After the operation, the Type 1 [2] transfer recipient is Jim") and a statement about person continuity ("After the operation, the Type 1 [2] transfer recipient is a person"). The ordinal position of the scenarios in the booklets was balanced across participants, and the questions appeared in a counterbalanced order, along with other questions about whether the post-op creature still had the same occupation and gender. The participants made their ratings by circling a number on a 0-to-9 scale, whose endpoints were labeled "strongly disagree" (= 0) and "strongly agree" (= 9).

A second group of participants received similar scenarios, but these stories differed in one important respect. Instead of the character's brain being transplanted to a new recipient, the character's

memories were copied onto a computer, and the computer was then placed into a robot or a spare human body. For example, corresponding to the description of a "Type 1 transfer" in Table 1a, the new scenarios contained these sentences: "In a 'Type 1 transfer procedure,' a team of doctors copies the memories in Jim's brain and transfers them onto a state-of-the-art computer. The computer is placed in a highly sophisticated cybernetic body (robot). Jim's original body is destroyed in the operation." Similarly, corresponding to the "Type 2 transfer" in Table 1b, participants read, "In a 'Type 2 transfer procedure,' a team of doctors copies the memories in Jim's brain and transfers them onto a state-of-the-art computer. The computer is placed in a stock body. Jim's original body is destroyed in the operation."

There were 64 participants in the experiment, 33 in the brain transplant group and 31 in the memory transfer group. All participants were Introductory Psychology students who took part in order to fulfill a course requirement.

Results and discussion. Participants gave very different agreement ratings to the statement that the post-op creature was still Jim than to the statement that it was still a person. In determining whether the creature was still Jim, participants paid special attention to whether Jim's memories survived and paid somewhat less attention to whether the recipient of these memories had a robot or a human shape. In the brain-transplant condition, for example, when Jim's memories remained intact after the operation, participants gave a mean agreement rating of 6.6 (on the 0-to-9 scale) to the statement that the recipient was still Jim, but a rating of only 2.0 when Jim's memories were not the same. The difference due to the human versus robot form of the product was much smaller. The mean rating for the human embodiment was 5.0 and for the robot embodiment 3.6. By contrast, when deciding whether the post-operative being was still a person, participants leaned much more heavily on whether it had the form of a robot or of a human than whether it preserved Jim's memories. The mean agreement rating for the statement that the recipient was still a person was 8.0 when the creature had a human body but 2.6 when it

had a robot body. The means for the memory contrast, however, were 5.7 (memory intact) and 4.9 (memory not intact).

The result of these different response patterns was that in some conditions participants were more likely to agree that the creature was still Jim than that it was still a person; in others, they were more likely to agree that the creature was still a person than that it was still Jim. These effects from the brain-transplant condition appear in Figure 1, where the open circles represent the question about Jim and the filled circles the question about personhood. Error bars indicate one standard error of the mean. (Results from the conditions in which Jim's memories were downloaded to a computer produced a pattern similar to that of Figure 1, but with a lower overall mean.) The important condition is the one in which the resulting creature has the body of a robot but the memories of Jim. In this case, participants are more apt to agree that the creature is Jim than that the creature is a person. Conversely, when the creature has the body of a human but memories that differ from Jim's, then participants are more likely to agree that the creature is a person than that it is Jim. This double dissociation presents difficulties for psychosortalism, since this theory predicts that Jim's existence should come to a halt when he stops being a person. We find, however, that Jim continues to exist, though "out of sorts."[5]

Please insert Figure 1 about here.

In our participants' view, then, Jim can outlive his personhood when his memories are intact in a robot. This finding is at odds with psychosortalism on the assumption that Jim's sortal is *person*, since Jim's membership in the person category should determine how long he persists as an individual. (See also Blok, Newman, Behr, & Rips, 2001, Experiment 2, for similar results involving scenarios like that in Table 1a). The following experiment generalizes our findings to individuals other than people.

Experiment 2: Transformations of Natural Kinds and Artifacts

To see whether individual natural kinds and individual artifacts can persist across changes in sortal categories, we used a variation on the procedure of Experiment 1. On each trial of the present study, we gave participants a picture and a short description for each of a set of common objects (e.g., a particular cat or cup), including a basic level sortal term (e.g., *cat* or *cup*) and a proper name for each item. For example, we told participants about "Jim's cat Bob. He is 2 years old and has a white coat. Jim inherited the cat from a former roommate. The cat is friendly but does not like it when anyone pets his belly." The participants also saw a drawing of a sci-fi device: either a "transporter" that we described as transporting the object particle-by-particle to a new place and reassembling it, or a "copier" that made a new copy of the object (while the old object was destroyed by a "disrupter ray"). Figure 2 shows pictures of these two devices, as we presented them to the participants. A picture of the object (e.g., a picture of Bob) appeared in the input part of the device, as in Figure 2. The participants then saw a picture of the outcome of the transformation in the assembler portion of the device (the original picture disappeared), and it was either the same picture as before, a picture of a related object, or a picture of an unrelated one. In Bob's case, the after-picture either was exactly the same as the before-(cat)-picture, was a picture of a dog, or was a picture of a boat. We then asked participants to judge whether the outcome of the transformation was still the same individual (still Bob?) and whether it was still a member of the same sortal (still a cat?).

Please insert Figure 2 about here.

We assume that participants will believe that the outcome of the transformation is no longer a member of the original sortal when the transformation produces something that looks like a member of a distinct category (related or unrelated). If the transformed object comes out looking like a dog or a boat, participants will judge it no longer a cat. (People know that appearances can be deceiving when it comes to category membership, as many previous transformation experiments demonstrate; but in the absence of information to the contrary, appearance provides evidence of category status. See Medin & Ortony,

1989). The more interesting question is whether participants will take the new object to be the same individual. Since the same particles (or a copy of them) go into the new object and since there are no other obvious contenders for the continuation of the old object, participants may be more likely to see the new object as retaining its identity than as retaining its sortal status. For example, participants may judge that if the transformed object looks like a dog or a boat, it may still be Bob, though no longer a cat, contrary to psychosortalism. Bob's persistence should be more likely when the outcome object is related than unrelated, since it's easier to imagine Bob surviving as a dog than as a boat.

Method. To select the stimulus objects for this experiment, we conducted a preliminary study in which we asked participants to identify pictures of common objects. Sortalists contend that sortals are the terms that answer the question *What is it?* for objects (Wiggins, 1980, 1997); so we can assume that if participants uniformly volunteer a particular term in answer to this question, then that term is likely to be the item's sortal. Participants in the preliminary study received a 48 page booklet, each page containing a 1.5" x 1.5" black and white photo of a different item (e.g., car, toaster, lizard, cat). We asked the participant to write down an answer to the question *What is it?* for the pictured item in a blank provided on the same page. The participant also rated the complexity of the pictured object on a 0-9 scale, where 0 was labeled "not complex" and 9 was labeled "very complex." Twenty participants took part in the preliminary test, and each participant received a different random order of the item*s*.

We discarded objects for which more than 5% of participants produced a category that disagreed with the most commonly mentioned one, and from the remaining objects, we selected 20 items to serve as the beginning state of the transformations in the main experiment. We refer to these items as *original objects* for this reason. The original objects included 10 natural kinds (apple, plant, leaf, pineapple, tree, cat, robin, turtle, person, and mouse) and 10 artifacts (fire hydrant, hammer, comb, chair, cup, sewing machine, toaster, refrigerator, car, and house). We further divided each of these two groups into *simple* and *complex* items, on the basis of the participants' complexity ratings in the preliminary study. The first five items in the parenthesized lists just mentioned were the simple natural kinds and artifacts; the last

five were the complex items. We also selected an additional 20 items (again, 10 natural kinds and 10 artifacts) to be the final states of the transformations. For the main experiment, we arranged the selected objects into triples, each consisting of an original object, a *related outcome object*, and an *unrelated outcome object*. If the original object was a natural kind, the related object was a (different) natural kind and the unrelated object was an artifact. If the original object was an artifact, the related object was a (different) artifact and the unrelated object was a natural kind. For example, the picture of the cat was grouped with a picture of a dog (related object) and a picture of a boat (unrelated object); the picture of the cup was grouped with a picture of a strainer (related object) and a rose (unrelated object). Items in each triple were matched for complexity as far as possible, based on the participants' complexity ratings.

At the beginning of the main experiment, we told participants that they would be making judgments about a hypothetical device. The instructions informed one group of participants about the transporter (see Figure 2a): An object is placed in the part of the device called a "splitter" and broken down into particles that are then sent through a particle pipeline. The particles arrive at an "assembler" which reconstitutes the particles into an object. A second group of participants read about the copier (see Figure 2b): In this case, the object is placed in a "scanner," and a replica of the object is then formed in the "assembler," while a "disrupter ray" destroys the original object.

Participants then received 20 computer-controlled trials, each of which described a transformation that the device carried out on one of the original objects. Each trial consisted of an initial screen that showed one of the original objects in the splitter or scanner of the device (as in Figure 2a or 2b) and two follow up screens. The initial screen also contained a brief paragraph that described the object as a member of a particular category (e.g., cat or cup) and gave its proper name, as in the information about Bob the cat that we quoted earlier. (Most of our objects are not the kinds of things that customarily have proper names, but the paragraphs specified nicknames that their owners had given them. E.g., the cup's owner names it "Jane.") The category in the description was always the one that we had obtained from the norms of the preliminary study. A second screen then revealed the results of the transformation—a picture of the outcome object in the device's assembler and an empty space in the splitter or scanner

where the original object had been. We provided no descriptive information about the outcome object, other than the picture itself. For a third of the participants, the outcome object was always the same as the original object (i.e., the same picture appeared in the assembler as had been in the splitter or scanner). A second group saw the corresponding related pictures as the outcome objects on all trials. For example, if the original object was a cat, the outcome object was a dog, and if the original object was a cup, the outcome object was a strainer. A final group saw the corresponding unrelated pictures as outcome objects on all trials. Thus, if the original picture was a cat, these participants saw a boat as the outcome, and if the original picture was a cup, they saw a rose.

A third screen contained two questions about the transformation. One question asked participants to rate their agreement with the statement *After the transformation, the object in the assembler is a* _____, where the blank was filled with the category name (e.g., *cat* or *cup*) that we had used to describe the original object. Participants made their rating on a 0-to-9 scale, where 0 was labeled "strongly disagree" and 9 was labeled "strongly agree." The second question asked participants to rate agreement with the statement *After the transformation, the object in the assembler is* _____, where the blank was filled with the proper name (e.g., *Bob* or *Jane*) for the original object. The computer randomized the order of the category continuity question and the individual continuity question on each trial.

The computer also randomly permuted the order of the 20 trials in a new way for each participant. During a trial, the participants could move from one screen to the next by pressing any key on their keyboard. They made their agreement ratings by pressing a key on the top row of the keyboard. The entire experimental session lasted a total of 20 to 45 minutes.

There were six conditions in this experiment, formed by combining the two device types (transporter or copier) and the three types of outcome (same, related, or unrelated outcome object). We randomly assigned six participants to each of these conditions. Participants were 36 students from an introductory psychology class, who took part in the study for course credit. These students were from the same pool as those in the preliminary experiment and in Experiment 1, but had not been in those earlier

studies. A computer failure lost the data from one of the participants (from the transporter condition with same outcomes). The results we analyze in the next section come from the remaining 35 participants.

Results and discussion. The main issue in this experiment is whether participants would see the individual artifacts and natural kinds as outliving their sortals. If the object that is input to the transporter or copier in Figure 2 is Bob the cat and the outcome object looks like a dog, are participants more likely to judge that the outcome object is Bob than that it is a cat? The results of the study provide evidence that this is so for related combinations like cat-dog and cup-strainer.

Figure 3 shows the mean ratings that are most relevant to this issue[6]. When the outcome object was the same as the original, participants judged the outcome to be a member of the same category but were less convinced that it was the very same individual. For example, even when the original and outcome objects were the same picture of a cat, participants were more likely to agree that the outcome was a cat than that it was Bob. Transporting or copying the objects' particles to a new place apparently threatens the objects' identity but not their category status. For transformations to related objects, however, the judgments reverse. Although participants' agreement ratings are lower overall, they are nevertheless more likely to agree that the outcome object is the same individual than that it is a member of the same category. When the original object is a cat picture and the outcome object a dog picture, for example, participants' agreement ratings are higher for the statement that the outcome is still Bob than that it is still a cat. This reversal is contrary to psychosortalism, as we have described it earlier. Finally, when the outcome object is unrelated to the original object, the agreement ratings are low and approximately equal for the identity and category questions. When the input is a cat picture and the output a boat picture, participants believe the thing is neither Bob nor a cat.[7]

Please insert Figure 3 about here.

A possible objection, which applies to both experiments, is that our sci-fi task may have been too unnatural to capture participants' "genuine" beliefs about object identity. Safe to say, our undergraduate participants do not encounter in their everyday routines the sorts of grotesque transformations that we

asked them to reason about in these experiments. Perhaps these unfamiliar and unrealistic scenarios led them to err in judgment; perhaps in a more natural terrain, their decisions would conform more closely to those of psychosortalists. We have already tried to head off a related objection in distinguishing sortalism from psychosortalism: Our concern here is not with expert metaphysical theories that derive from careful analysis, but with more immediate intuitions from novice judges. Still, is it reasonable to study intuitions in a context so far removed from the mundane?

Our objective in using the transformation paradigm was to investigate what is necessarily true of object identity, not just in what is customarily true; so some departure from business-as-usual is essential in order to check our hypotheses at all. Moreover, although we hope our participants have not experienced such transformations first hand, we believe that speculation about the effects of similar changes is quite natural and familiar. Closely related scenarios figure prominently in fiction and even in non-fiction. The cover of the most recent issue of *Scientific American* taunts, "Teleportation is simple. Ready for a real challenge?" Our participants could hardly avoid some acquaintance with such scenarios, and none of our participants complained of being unable to make sense of them.

Sorting Things Out

We've documented several cases in these experiments where an individual appears to survive its own sortal. In Experiment 1 if John's memories were programmed into a robot body, participants were more likely to agree the result is John than that the result is a person. In Experiment 2 if the transformed object had the outward appearance of a related natural kind or artifact, participants were more likely to agree that it was the same individual than that it was a member of the original sortal. For instance, when Bob the cat came out of the assembler looking for all the world like a dog, participants thought him more likely to be Bob than to be a cat.

These results seem to be bad news for psychosortalism, since according to this doctrine people think that the continuity and the continued existence of individuals depend on their sortal. Before saying goodbye to psychosortalism, though, we need to consider some possible variations that might get this

theory out of trouble. The most obvious of these is that we simply picked the wrong categories. Although John can survive as the same individual after becoming a nonperson, maybe that's because John's sortal isn't *person* but something else. Ditto for Bob the cat. Maybe Bob's sortal is something other than *cat*. One initial hurtle for this objection is that we took pains in the preliminary study for Experiment 2 to elicit for each pictured object the name of the category that answered the question *What is it?* Since this is the sortal category according to sortalist theories (Wiggins, 1980, 1997), it isn't likely that we overlooked the correct category.

Physical Object as a Sortal

Another possibility for psychosortalists, however, is to suppose that an object can have more than one sortal and therefore more than one set of identity conditions. Although Bob's lower-level sortal is *cat*, perhaps he retains a higher-level sortal that can sustain his existence while he is having out-of-cat experiences. And here there is an actual proposal to consider, thanks to developmental psychologists (e.g., Carey, 1995; Xu, 1997). The idea is that the category of physical object—"a three dimensional, bounded entity that moves on a spatiotemporally continuous path"—can function as a sortal to guide tracking of individuals when more specific sortals are not at hand. According to this view, infants lack lower-level sortals like *cat* or *cup* to help them recognize solid objects' identity but instead use the concept of a physical object to follow perceived individuals in the environment. The same concept has been used to explain transformation cases somewhat similar to ours, such as how readers of the book of Genesis can understand Lot's wife's continued existence as a pillar of salt. Perhaps in a similar way, participants in our experiment thought that Bob the cat and John the person continue to exist (even when no longer cat or person) because they remain physical objects.

Xu's (1997) defense of *physical object* as a sortal is a valiant one (see the rejoinders by Ayers, 1997; Hirsch, 1997; and Wiggins, 1997), but we doubt that it is the right explanation for our findings. For one thing, the transformations in Experiments 1 and 2 did not necessarily respect the criteria for physical object that we quoted from Xu in the preceding paragraph. Recall that the scenarios in Experiment 1

specified that John's physical body is destroyed when his brain is transplanted to a new human or robot body. In one condition, John's brain is transplanted (see Table 1); in another, his brain is also destroyed and only his memories are downloaded to their new home. The result of the operation does not seem to be the same physical object as before (in the technical sense of same "three-dimensional bounded entity that moves on a spatiotemporally continuous path"), but participants nevertheless identified it as John. If participants are using a sortal to support their judgment, it can't be *physical object*. Although brain transplants produced overall higher ratings than memory downloads, even in the download condition participants seem willing to agree that the post-op result is still John when memories are preserved. What mostly seems effective in preserving John is the Lockean continuity of his memories and not the continuity of his body. The same conclusion follows from the results of the copier condition in Experiment 2.

Second, as Hirsch (1997) and Wiggins (1997) have pointed out, the notion of physical object that figures in Xu's proposal excludes things like trees, fire hydrants, houses, and many other objects that don't move and that aren't customarily detached from their surroundings. This could merely indicate that the concept in question isn't a faithful analysis of adults' everyday physical-object concept (assuming that people regard trees, houses, etc., as bona fide physical objects). But if we regard "physical object" as a technical term in this context that excludes trees and other stationary items, it then has trouble explaining why these items exhibited the same behavior in Experiment 2 as moveables such as cars and cats. We deliberately included a fire hydrant, a tree, and a house among the original objects for this reason, and the means for these items closely match those for the whole data set in Figure 3.[8] This again suggests that whatever is responsible for judgments of identity continuity in these experiments isn't the physical object concept.

Third, if what is preserving the identity of the Experiment 2 individuals is *physical object*, then there is no clear reason why identity judgments should decline between related and unrelated transformations. In both cases, the outcome of the transformation remains a physical object (omitting the stationary items just discussed) and is no longer a member of the lower-level sortal category. Hence,

there is no explanation for why identity is judged more likely to be preserved across related than unrelated categories.[9]

Explanations of Other Sorts

Of course, psychosortalists have lots of room to maneuver between cats and physical objects. Perhaps some of the participants in Experiment 2 took Bob's sortal to be animal rather than cat, despite the results of our preliminary experiment and our assertion in the main experiment that Bob was a cat. In that case, when the outcome object happens to look like a dog, Bob's sortal is preserved and so is Bob's identity. Thus, superordinate sortals appear to do a better job at explaining these results than do basic level or physical object concepts. We're still left with puzzles, however. What superordinate sortal mediates John's continuity as a robot? Sentient being? What explains the fact that when the outcome object looks the same as the original object in Experiment 2 participants are in better agreement that the thing is a cat than that it is Bob?

Still, it is worth considering the possibility that people have multiple sortals at different hierarchical levels. It is clearly not possible for all predicates to function as sortals without giving up sortalists' key distinction between sortal and nonsortal predicates. Also, as we have just seen, very general sortals, such as *physical object*, are unlikely to explain the data. But what about the possibility that both basic-level categories like cat and superordinate categories like animal provide identity criteria for Bob? Some sortalist theories seem compatible with the idea of rival identity principles (Geach, 1962), and this possibility seems consistent with the form of the results in Figure 3. But as other sortalists have pointed out (Wiggins, 1980), it is difficult to formulate this multiple sortal notion consistently. What does it mean to say, for example, that a basic-level sortal like *cat* determines Bob's persistence if he can be resurrected simply by virtue of his dual membership in the superordinate *animal*? To be sure, you can say that Bob ceases to exist as a cat when he is no longer a cat. But this is equally true of all predicates, sortal and nonsortal: Bob ceases to exist as dirty when he is given his bath. Attempts to set up multiple sortals at different taxonomic levels seem to deprive the lower-level categories of their sortal status.

Nonsortalist Explanations

We propose that psychosortalism is handicapped in explaining how people trace object identity because people use a different source of information for this purpose. As we have already suggested, we think people are realists about individual natural kinds and artifacts in the sense of believing that external causal forces launch these individuals, support them during their careers, and finish them off in the end. It's usually the continuity of these forces that determines the continuity of the individuals, not the particular categories that the individuals happen to belong to. Thus, if causal conditions are right, John can be repackaged as a robot and Bob as a dog. At this point, though, psychosortalists have a predictable comeback that goes like this: "Sure, causal factors are important in determining individuals, but what kind of causal factors are you talking about? If causal factors are crucial, then according to your own experiments, it appears that what's important in insuring John's survival is the causal process that preserves John's memories, what's important in insuring Bob's survival is the causal process that preserves Bob's biological properties, and what's important in insuring Jane the Cup's survival is the causal process that preserves Jane's physical or material properties. You yourself have just argued that these processes can't be domain independent. It really looks as though the effective causal processes vary with the sort of object at issue, which is what we psychosortalists have been saying all along. So talk of causal properties merely begs the question against psychosortalists; it doesn't provide an alternative theory."

It's common ground that different causes are responsible for the survival of different individuals. An individual person and an individual french fry depend on different physical laws to maintain their existence and identity. This implies that, given sufficient flexibility about what counts as a category (given ad hoc categories in which any set of elements is a category), we could partition individuals into categories that have like individuating conditions. If we call these partitions "sortals," then sortals are perfectly correlated with the individuating processes. But a psychosortalism of this kind is vacuous. It's obvious in advance that we can partition objects into arbitrary groups that have the same identity

conditions. Sortalism and psychosortalism are nontrivial claims because they assert that sortals are (metaphysically or psychologically) *prior* to the individuals they carve out. According to this top-down view, people have no conception of an individual before they've acquired an appropriate sortal. Sortals are therefore not generalizations that people discover on the basis of their experience with individuals or individuation. This direction of conceptual priority also places limits on psychosortalists' freedom to pick and choose categories to serve as sortals in their theories. The game plan for psychosortalists is to specify sortals that could plausibly have this prior conceptual role and then to show that people's judgments about individuals are shaped by these sortals. What the experiments in this article imply, however, is that those categories that one might have thought have the best claim to cognitive priority—essentially basic-level categories that people use to answer *What is it?*—do not set the boundaries for the individuals that fall under them. We suspect this is because a purely top-down approach to individuation doesn't match the methods people ordinarily employ.

A related objection to our antipsychosortalist viewpoint is that causal relations seem to presuppose individuated objects—they presuppose a cause and an effect. Thus, causal relations can't be responsible for the very same objects. However, it is not necessarily the case that causes and effects are objects of the type whose individuation we are trying to explain. According to some theories of causation (e.g., Davidson, 1967), causes and effects are events—particular activities or accomplishments. Certainly, *some* causes are events (e.g., mutations in a protein or activities on an assembly line), and these may be all we need to explain individuation of the natural kinds and artifacts in question.

It's true, however, that we don't have on offer a detailed theory of people's beliefs concerning the individuation of specific entities. What we have are hints about the factors responsible for continuity in particular cases we have studied. With respect to persons, for example, Experiment 1 suggests that our participants attended to maintenance of memories, but also maintenance of substance, in deciding whether Transformed John is John. If so, commonsense beliefs about personal identity seem to be neither wholly physicalist nor wholly functionalist, but some mixture of these (see also Blok et al., 2001). It is also consistent with our position that there may be no closed set of factors underlying identity, but instead an

open set whose membership depends on the range of possible continuations for a given individual—contenders for the future being of the individual—as Nozick (1981) suggested. The project of mapping this decision strategy, though, is one that goes beyond the scope of this paper.

Concluding Comments

Some early models of categorization posited that people could do without information specific to categories, relying instead on information about exemplars of these categories and making inductive inferences when necessary to the entire set. The mental representation of cats, according to this story, consists of memories of individual cats you've met, with no special summary information about cats as a group. These exemplar models did well in accounting for the data from experiments in which participants viewed small groups of novel exemplars (e.g., two groups of schematic faces) and then decided to which group each of a set of transfer exemplars belonged. However, exemplar models encountered difficulties in explaining commonsense judgments about everyday categories. To say that dodos are extinct, for example, doesn't seem equivalent to saying something about individual dodos. *Extinct* isn't the sort of predicate that applies to individuals, much less to ones you've met (Krifka, Pelletier, Carlson, ter Meulen, Link, & Chierchia, 1995; see Rips, 1995, for a discussion of other problems with exemplar theories). Likewise, the results of the experiments we've reported here suggest that "exemplars" can sometimes free themselves from the bounds of even basic level categories, with the same exemplar being now a cat and now a dog. If so, then it is hard to see how knowledge of cats could be coextensive with knowledge of exemplars.

Psychosortalism is another attempt at specifying the relation between individual and category concepts, but from the opposite direction. On this approach, people have no representation of individuals apart from their representation as members of a category. "No individuation without classification" is the motto. There is no representation of Bob, for example, except as a member of the cat kind. This theory does not seem committed to the semantic gaffs of exemplar models, and sortalism has deep insights into the space of possible ways that stuff could be individuated. Despite its greater sophistication, though,

psychosortalism, like exemplarism, seems to us to overstate the dependence of individual and category concepts. The category-hopping individuals in our studies are evidence that people don't believe that knowledge of categories *reduces* to knowledge of exemplars. But they're equally evidence against the idea that knowledge of categories *dictates* knowledge of exemplars.

References

Ayers, M. (1997). Is *physical object* a sortal concept? A reply to Xu. *Mind & Language, 12*, 393-405.

Blok, S., Newman, G., Behr, J., & Rips, L. J. (2001). Inferences about personal identity. *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, 80-85.

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

Carey, S. (1995). Continuity and discontinuity in cognitive development. In D. N. Osherson (Series Ed.) & E. E. Smith & D. N. Osherson (Vol. eds.), *An invitation to cognitive science: Vol. 3 Thinking* (2nd ed., pp. 101-129). Cambridge, MA: MIT Press.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language, 16*, 37-55.

Davidson, D. (1967). Causal relations. *Journal of Philosophy, 64*, 691-703.

Dummett, M. (1981). *Frege: Philosophy of language* (2nd ed.). Cambridge, MA: Harvard University Press.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33-66.

Geach, P. T. (1962). *Reference and generality* (emended ed.). Ithaca, NY: Cornell University Press.

Gelman, S. A., & Hirschfeld, L. A. (1999). How biological is essentialism? In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 403-446). Cambridge, MA: MIT Press.

Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Zoology, 23*, 536-544.

Gottfried, G. M., Gelman, S. A., & Schultz, J. (1999). Children's understanding of the brain: From early essentialism to biological theory. *Cognitive Development, 14*, 147-174.

Grene, M. (1963). *A portrait of Aristotle.* London: Faber and Faber.

Gupta, A. (1980). *The logic of common nouns: An investigation in quantified modal logic*. New Haven: Yale University Press.

Gutheil, G., & Rosengren, K. S. (1996). A rose by any other name: Preschoolers understanding of individual identity across name and appearance changes. *British Journal of Developmental Psychology, 14,* 477-498.

Hall, D. G. (1998). Continuity and persistence of objects. *Cognitive Psychology, 37,* 28-59.

Hirsch, E. (1997). Basic objects: A reply to Xu. *Mind & Language, 12,* 406-412.

Hobbes, T. (1839-1845). De corpore. In W. Molesworth (Ed.), *The English works of Thomas Hobbes* (Vol. 1). London: John Bohn.

Hull, D. (1978). A matter of individuality. *Philosophy of Science, 45,* 335-360.

Johnson, C. N. (1990). If you had my brain, where would I be? Children's understanding of the brain and identity. *Child Development, 61,* 962-972.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175-219.

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

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., & Chierchia, G. (1995). Genericity: An introduction. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1-124). Chicago: University of Chicago Press.

Kripke, S. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Kruschke, J. K. (1992). ALCOVE: An exemplar based connectionist model of category learning. *Psychological Review, 99,* 22-44.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

Littschwager, J. C. (1995). Children's reasoning about identity across transformations. *Dissertation Abstracts International, 55* (10), 4623B. (UMI No. 9508399).

Locke, J. (1694/1979). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Oxford: Clarendon Press. (Original work published 1694).

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review,109,* 376-400.

Loux, M. J. (1991). *Primary ousia: An essay on Aristotle's Metaphysics Z and H.* Ithaca, NY: Cornell University Press.

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

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

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

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

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression.* Reading, MA: Addison Wesley.

Nosofsky, R. M. (1986). Attention, similarity, and the indentification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nozick, R. (1981). *Philosophical explanations.* Cambridge, MA: Harvard University Press.

Parfit, D. (1984). *Reasons and persons.* Oxford, England: Oxford University Press.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure. Typological studies in language, vol. 45.* (pp. 137-157*).* Amsterdam: John Benjamins.

Quine, W. V. (1960). *Word and object.* Cambridge, MA: MIT Press.

Quine, W. V. (1969). Ontological relativity. In *Ontological relativity and other essays* (pp. 26-68). New York: Columbia University Press.

Rips, L. J. (1995). The current status of research on concept combination. *Mind & Language, 10,* 72-104.

Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin, 127,* 827-852.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Shoemaker, S. (1970). Persons and their past. *American Philosophical Quarterly, 7,* 269-285.

Smith, E. R, & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3-21.

Wiggins, D. (1980). *Sameness and substance.* Cambridge, MA: Harvard University Press.

Wiggins, D. (1997). Sortal concepts: A reply to Xu. *Mind & Language, 12,* 413-421.

Wilcox, T., & Baillargeon, R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37,* 97-155.

Xu, F. (1997). From Lot's wife to a pillar of salt: Evidence that *physical object* is a sortal concept. *Mind & Language, 12,* 365-392.

Xu, F. (in press). The development of objection individuation in infancy. In J. Fagen & H. Hayne (Eds.), *Progress in infancy research* (vol. 3). Mahwah, NJ: Erlbaum.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30,* 111-153.

Authors' Note

We thank Jennifer Asmuth, Dan Bartels, Jennifer Behr, Amber Bloomfield, Douglas Medin, Ariela Lazar, Beth Lynch, Jeff Pasch, Andrea Proctor, Eyal Sagi, Russ Burnett, and Elizabeth Spelke for their help on an earlier version of this paper. Thanks to Jeff Rice for the cute cat name. NSF Grant SES-9907414 supported this research. Correspondence about this paper should be sent to Lance Rips, Psychology Department, Northwestern University, Evanston, IL 60208. Email: rips@northwestern.edu

Table 1
Scenarios for Person Transformations, Experiment 1.

a. Brain transplant to robot recipient:

Jim is an accountant living in Chicago. One day, he is severely injured in a tragic car accident. His only chance for survival is participation in an advanced medical experiment called a "Type 1 transplant" procedure. Jim agrees.

It is the year 2020 and scientists have developed incredibly sophisticated computers and robots. In a "Type 1 transplant procedure," a team of doctors removes Jim's brain and carefully places it in a highly sophisticated cybernetic body (robot). Jim's original body is destroyed in the operation.

After the operation, all the right connections between the robot and the brain have been made, and the brain is able to control the robot. When the doctors turn on the robot, the robot appears to be human-like in its behavior. It has senses and can move and talk. The doctors scan the brain of the transplant recipient and note that the memories in it [NO memories in it] are the same as those that were in the brain before the operation. [Something must have happened during the transplant.]

b. Brain transplant to humanoid recipient:

Jim is an accountant living in Chicago. One day, he is severely injured in a tragic car accident. His only chance for survival is participation in an advanced medical experiment called a "Type 2 transplant" procedure. Jim agrees.

It is the year 2020 and scientists are able to grow all parts of the human body, except for the brain. A stock of bodies is kept cryogenically frozen to be used as spare parts in the event of an emergency. In a "Type 2 transplant procedure," a team of doctors removes Jim's brain and carefully places it in a stock body. Jim's original body is destroyed in the operation.

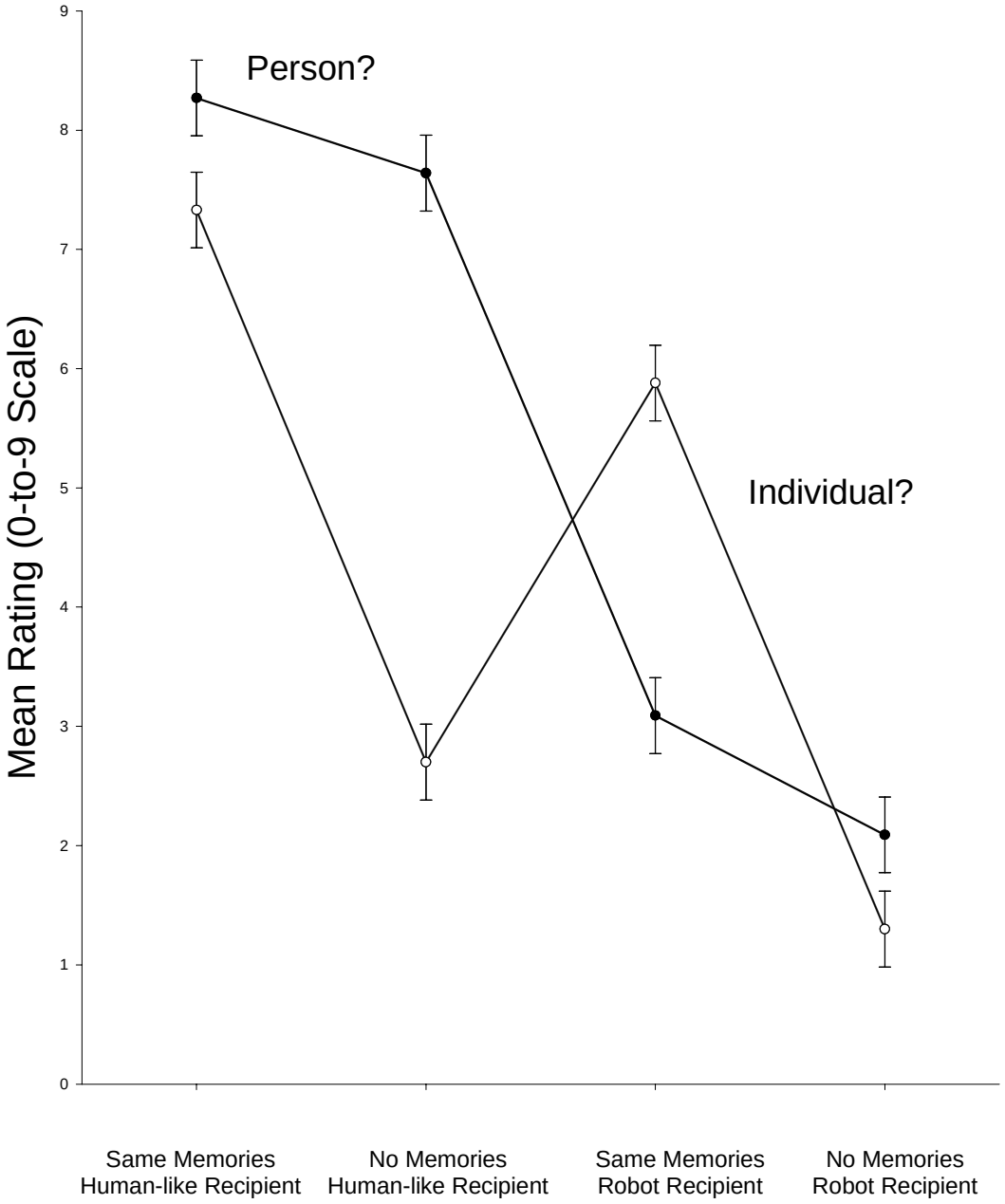
After the operation, all the right neural connections between the brain and the body have been made. The doctors test all physiological responses and determine that the transplant recipient is alive and functioning. The doctors scan the brain of the transplant recipient and note that the memories in it [NO memories in it] are the same as those that were in the brain before the operation. [Something must have happened during the transplant.]

Figure Captions

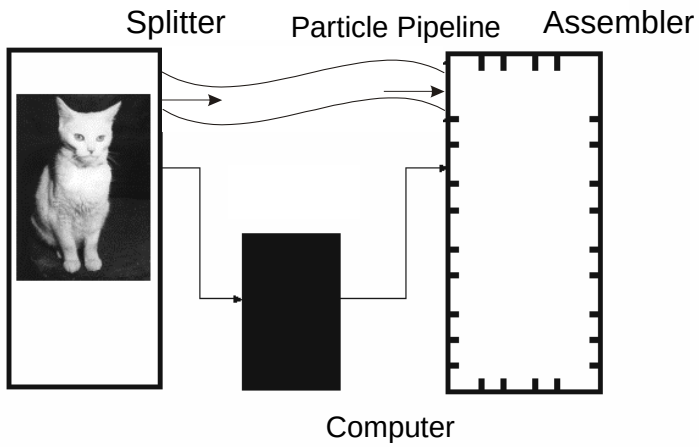
Figure 1. Mean agreement ratings (0-to-9 scale) for the statements that the transfer recipient is "still Jim" (open circles) and "still a person" (filled circles). Results are from brain transplant condition. The x-axis represents the four versions of the Experiment 1 story (see Table 1). The error bars indicate one standard error of the mean, based on 64 observations per point.

Figure 2. The two "devices" that performed the transformations in Experiment 2. (a) The "transporter" dissolves the object in its splitter, sends the particles through the pipeline, and reassembles an object in the assembler. (b) The "copier" copies the object in its scanner, reassembles an object in the assembler, and destroys the original object.

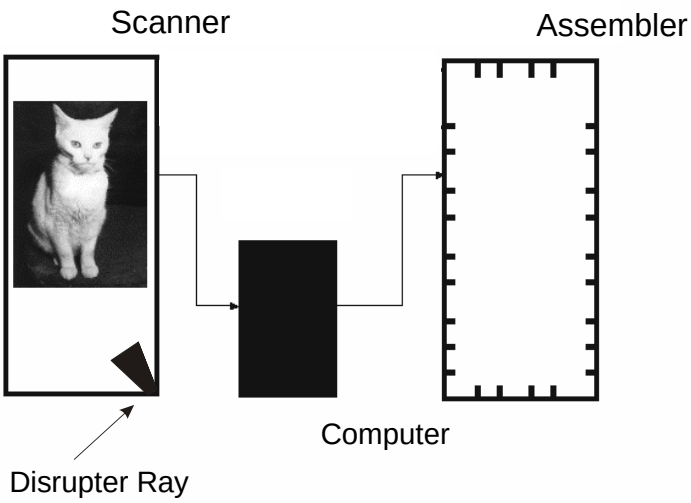
Figure 3. Mean agreement ratings (0-to-9 scale) for the statements that the outcome object is still the same individual (open circles) and still a member of the same category (filled circles). The x-axis represents the three types of transformation (original and outcome objects shown as the same picture, related pictures, or unrelated pictures). The error bars indicate one standard error of the mean.

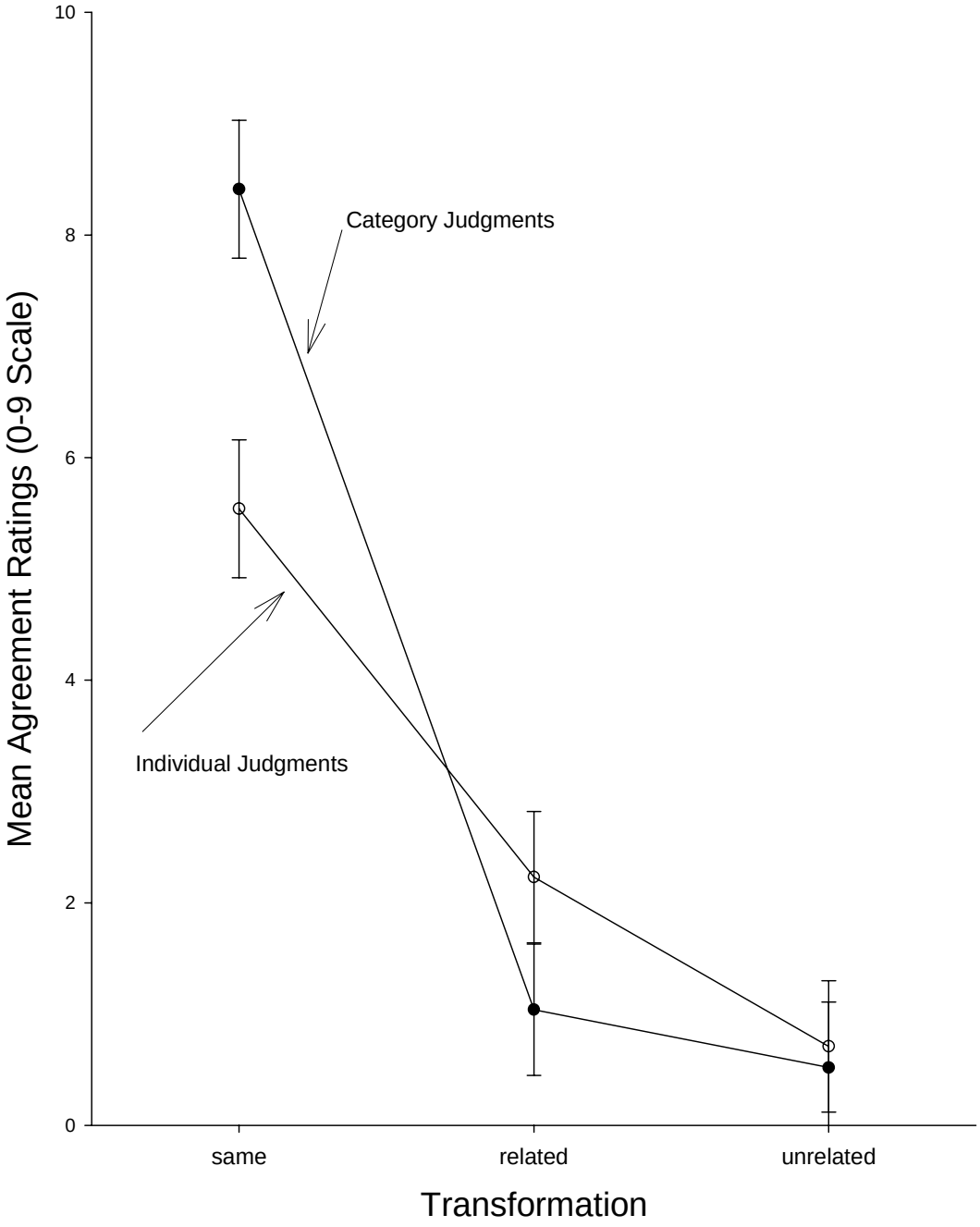

Person?
Individual?
Mean Rating (0-to-9 Scale)
0
1
2
3
4
5
6
7
8
9
Same Memories
Human-like Recipient
No Memories
Human-like Recipient
Same Memories
Robot Recipient
No Memories
Robot Recipient

a. Transporter



b. Copier




10
8
6
4
2
0
Mean Agreement Ratings (0-9 Scale)
Category Judgments
Individual Judgments
same
related
unrelated
Transformation

Footnotes

[1] For the ancient version, see Grene (1963) and Loux (1991) on Aristotle's *Metaphysics*. For the modern versions, see the sources cited in the following paragraph.

[2] It may seem problematic at the outset that an individual can fall under a sortal like *sapling* at t_1 but not at t_2 (when it is a mature tree) and yet continue as the same individual. One sortalist response to this issue (Wiggins, 1980, pp. 24-27) distinguishes *substance sortals* (e.g., *tree*) that apply to individuals through their entire life span and *phased sortals* (e.g., *sapling*) that apply to individuals during some part of their life. Substance sortals provide privileged information about identity. Thus, *tree* yields the principle of identity which determines that a particular immature mulberry tree at t_1 is the same as a mature one at t_2. Phased sortals coincide with this verdict, however, given appropriate adjustment to the tense of the identity statement. The sapling at t_1 *will be* the same as the mature tree at t_2. In Section 5, we consider other cases in which putative sortals appear to overlap in their application.

[3] We use the terms *bottom-up* and *top-down* advisedly, since *bottom-up* in this context does not mean perceptual. Antisortalism need not contend that object individuation, enumeration, and persistence are purely perceptual matters, and as our reference to causal factors suggests, we think it likely that nonperceptual beliefs about mechanisms play a key role in the way (people think) objects are identified. The important distinction is whether object individuation depends on knowledge of the categories to which the object belongs (top-down) or, more directly, on knowledge of the exemplar itself (bottom-up).

[4] There are rival philosophical theories that do not tie individuation to sortals. For example, Quine's (1960, 1969) view of individuation makes it a matter of broader "analytical hypotheses" or "background theories," rather than something determined on a category-by-category basis. According to this view, sortal terms such as *rabbit* don't succeed in pinning down specific individuals, since there are alternative referents for these terms (the famous undetached rabbit part or rabbit time-slice) that are

equally consistent with the evidence (given compensating adjustments in the meaning of other predicates). Because of these incompatible standards, there is no fact of the matter about what qualifies as an individual. Individuals are only identifiable relative to theories in which they play a role (as values of the theories' variables). Thus, although Quine's position reject sortalism, it is even further removed from the bottom-up position that we sketched above. (Another way of stating the antipsychosortalist view is that people believe—perhaps incorrectly—that one of the rival Quinean background theories is true.)

[5] A mixed-model analysis of variance confirmed these observations. Continuity rating (for identity and personhood) was the dependent measure. The type of procedure (brain transplant vs. memory copy) served as a between-participant factor. The other three factors were within-participants: memory continuity (intact or disrupted), post-op body (robot vs. humanoid), and question type (whether it is still a person vs. whether it is still Jim). The ANOVA confirmed that there was a main effect of having a humanoid body recipient, $F(1,62) = 124.26$, $p < .0001$. This effect of having the human form was larger for the personhood question than for the question about Jim, $F(1,90) = 84.24$, $p < .001$. There was also a main effect of intact memory, $F(1,62) = 103.32$, $p < .0001$, and this effect was larger for the question about Jim than for the question about personhood, $F(1,85) = 79.08$, $p < .0001$. Planned comparisons between questions were reliable in both critical conditions. When the recipient had a human form but no memories were preserved, the personhood question received higher ratings than the question about Jim, $F(1,233) = 133.94$, $p < .0001$. The reverse was true when the recipient had a robot form but had intact memories—the question about Jim received higher ratings than the question about personhood, $F(1,233) = 42.23$, $p < .0001$.

[6] There were no differences due to the type of device that performed the transformation—no difference between the transporter and the copier in Figure 2—and no interactions of device with the other factors in the analysis.

[7] To confirm these findings, we used a mixed-model analysis similar to that in Experiment 1, with continuity ratings serving as the dependent measure. First, there was a significant effect of transformational distance: Agreement ratings decrease from same pictures to related picture and from related pictures to unrelated ones, $F(2,32) = 46.64$, $p < .0001$. Second, as the result of the cross-over in Figure 3, there is also an interaction with question type (identity vs. category continuity), $F(2, 32) = 9.13$, $p < .001$. Planned comparisons at each transformational distance show greater agreement with the category than the identity question when the pictures were the same ($F(1, 32) = 16.28$, $p < .001$), marginally greater agreement with the identity than the category question when the pictures were related ($F(1, 32) = 3.05$, $p = .09$), and no significant difference when the pictures were unrelated ($F < 1$). In addition, we analyzed continuity responses as a function of the type of category of the original object. There was no main effect of whether the original object was a natural kind or an artifact and no significant interactions of type of category with (identity vs. category) questions.

[8] For the three items in question (fire hydrant, tree, and house), the mean agreement ratings for identity continuity were 5.82 (same transformation), 2.06 (related transformation), and 0.69 (unrelated transformation). Means for category continuity were 8.39 (same), 0.61 (related), and 0.66 (unrelated). These can be compared to the overall means in Figure 3.

[9] Some psychologists (e.g., Carey, 2001; Feigenson, Carey, & Spelke, 2002) have recently suggested that many empirical results on infant individuation and enumeration are best explained by attentional mechanisms, such as object files (Kahneman, Treisman, & Gibbs, 1992). These infant findings are the same ones that originally prompted the psychosortalists to advance the claim that *physical object* is a sortal. One possibility is that lower-level attentional mechanisms remove the need for a *physical object* sortal. Another possibility is that object files somehow depend on a higher-level sortal. Sortal afficionados seem not to have resolved this issue.