

# Tracing the Identity of Objects

Lance J. Rips  
Northwestern University

Sergey Blok  
University of Texas at Austin

George Newman  
Yale University

This article considers how people judge the identity of objects (e.g., how people decide that a description of an object at one time,  $t_0$ , belongs to the same object as a description of it at another time,  $t_1$ ). The authors propose a causal continuer model for these judgments, based on an earlier theory by Nozick (1981). According to this model, the 2 descriptions belong to the same object if (a) the object at  $t_1$  is among those that are causally close enough to be genuine continuers of the original and (b) it is the closest of these close-enough contenders. A quantitative version of the model makes accurate predictions about judgments of which a pair of objects is identical to an original (Experiments 1 and 2). The model makes correct qualitative predictions about identity across radical disassembly (Experiment 1) as well as more ordinary transformations (Experiments 2 and 3).

*Keywords:* object concepts, concepts, identity, causality, exemplars

Near the winter holidays, many of us receive greeting cards and accompanying photocopied letters from friends whom we have not seen in years. The letters provide news, usually of vacation trips and children's successes, but occasionally of more important life changes, allowing us to keep track of these friends and update our knowledge of them. Our initial encounters with these friends may have given us a rich stock of perceptual information, and this information may survive as part of our representation. However, unless snapshots accompany these cards, we have to track our friends on the basis of nonperceptual facts. Our surviving images may be radically out of date (Bjork, 1978). Still, the letters may provide enough nonperceptual, descriptive information to allow us to reidentify these people—to determine who in 2006 is Aunt Florence, the same individual we last saw in 1970. At a minimum, higher level information about identity must come into play when perceptual information is absent. Although we sometimes misrec-

ognize people and other objects we know (Young, Hay, & Ellis, 1985), nevertheless we are often able to keep track of individuals across lapses of attention, sleep, and other perceptual interruptions. Even preschool children can follow individuals over changes in perceptual properties (e.g., Gutheil & Rosengren, 1996; Hall, Lee, & Bélanger, 2001; Hall, Waxman, Brédart, & Nicolay, 2003; Sorrentino, 2001). We can, therefore, meaningfully ask what sort of knowledge is relevant to such abilities.

The aim of the present article is to examine the nature of concepts that are rich enough to support this type of thinking. We use the term *singular concept* to denote a cognitive representation of a unique individual, and we contrast singular concepts with *general concepts*, which are representations of categories. A representation of the Sears Tower is a singular concept in these terms, but our representation of (the category of) buildings or skyscrapers is a general one. Our focus in this article is on the way singular concepts promote judgments of the identity of individuals: the identity of specific everyday physical objects, such as most natural objects and artifacts. We would like to know what people take to be the ultimate basis for this identity, and the route we take in addressing this issue examines how people trace identity over time—how they determine that an individual at one time and situation is the same individual at another time and situation. Unlike most earlier research on object identity, however, the situations in question are not limited to those we actually perceive. Thus: the central issue is

Question 1. Given knowledge about a target individual  $x_0$  in some situation  $S$ , how do we decide whether this individual continues to exist in another situation  $S'$ , and if so, which of the individuals  $x_1, x_2, \dots, x_n$  in  $S'$  is the same as  $x_0$ ?

We believe that Question 1 reveals central facts about singular concepts and identity. By comparing potential answers to Question 1 in an experimental context, we can begin to determine which factors are most crucial to our notion of the identity of things.

---

Lance J. Rips, Psychology Department, Northwestern University; Sergey Blok, Department of Psychology, University of Texas at Austin; George Newman, Department of Psychology, Yale University.

This work was supported by National Science Foundation Grant SES-9907414. We also thank the Fulbright Foundation and the Catholic University of Leuven for support during the time we wrote this article. We thank Jennifer Asmuth, Dan Bartels, Jennifer Behr, Amber Bloomfield, Aveen Farooq, Liepa Gust, Ariela Lazar, Beth Lynch, Barbara Malt, Douglas Medin, Jeff Pasch, Andrea Proctor, Eyal Sagi, Jeffrey Sherman, Steven Sloman, Elizabeth Spelke, and Fei Xu for their help on earlier versions of this article. We also thank audiences at Northwestern University, University of Michigan, Catholic University of Leuven, and Warwick University for their comments. Some of the ideas included here developed in classes on object identity at Northwestern University, and we thank the students in these classes for their many suggestions.

Correspondence concerning this article should be addressed to Lance J. Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. E-mail: rips@northwestern.edu

### Properties of Identity

In this context, asking whether individual  $x_i$  is the same as  $x_0$  means asking whether  $x_i$  is numerically identical to  $x_0$ . This is the equality relation that holds between each specific thing and itself,  $x_i = x_0$ . When we refer to objects being identical in what follows, it is to this relation of numerical identity or equality and not to the relation of looking alike or sharing many qualitative properties (as in identical twins). In deciding among theories of individual identity, it is helpful to keep the basic properties of this relation in mind. According to most treatments, individual identity is reflexive, symmetric, and transitive (e.g., Mendelson, 1964). That is, for any  $x_i$ ,  $x_j$ , and  $x_k$ :

Principle 1a.  $x_i = x_i$  (reflexivity).

Principle 1b. If  $x_i = x_j$ , then  $x_j = x_i$  (symmetry).

Principle 1c. If  $x_i = x_j$  and  $x_j = x_k$ , then  $x_i = x_k$  (transitivity).

It is possible that people's judgments of identity sometimes violate these principles, as we discuss later, but they provide a starting point for theory development. To specify the identity relation more precisely, we can add a principle that is also widely agreed to characterize numerical identity. This principle, sometimes called Leibniz's law, is that if two objects are identical, then any property true of one is also true of the other. This can be expressed as follows:

Principle 2. For any property  $F$ , if  $x_i = x_j$ , then  $Fx_i$  if and only if  $Fx_j$  (Leibniz's law).

Leibniz's law does not prohibit objects from changing properties across time. It guarantees only that if  $x_i = x_j$ , then if  $x_i$  has a particular property at a time,  $x_j$  must have the same property at that time.

Although Principles 1 and 2 may seem obvious, we will see shortly that it is not always easy to square them with people's thinking.

### Orientation and Overview

The literature on concepts in cognitive psychology centers mainly on general concepts, such as cats and buildings, rather than on concepts of individuals, such as a specific cat or a specific building. No one would deny that we have singular concepts, especially in the case of people (e.g., Woodrow Wilson, Aunt Florence), pets (e.g., our calico Cat-a-tonic), or other specially named objects (Sears Tower, Amazon River, Green Street, Guernica, and Macy's). These individuals are often important to us, and our concepts of them inform our beliefs about the uniqueness and worthiness of things. Singular concepts are of obvious interest to investigators in social, personality, and clinical psychology because of the centrality of the self concept (e.g., Kihlstrom & Cantor, 1984; Markus & Nurius, 1986) and the concepts of specific others (e.g., Klein, Loftus, Trafton, & Fuhrman, 1992; Park, 1986), in episodic and autobiographical memory because of the individuality of personal experiences (e.g., Conway, 1990), and in areas of psychology that focus on individual differences. However, research in the concepts-and-categories area implicitly views singular concepts as less problematic than general concepts, probably

because singular concepts often match the objects we perceive everyday. Although we can readily perceive Cat-a-tonic, we cannot perceive the cat species. In the same vein, there is hardly any doubt among psychologists that Cat-a-tonic exists in the real world, independent of our thought and language, but there may be real doubt about whether the cat species exists in a similar way, independent of people's classifying abilities.

Exemplar models of categorization (e.g., Medin & Schaffer, 1978) take advantage of this difference in status by proposing that people represent categories by means of singular concepts. According to this view, for example, people represent the category of cats as a set of individual cats that they have encountered and remembered. However, although the exemplar model may be the correct approach to general concepts, representing categories in terms of exemplars leaves many cognitive issues unresolved. As we have already noted, the representation of exemplars has to include conceptual as well as purely perceptual information to explain the way we trace identity across time; thus, reducing general concepts to singular concepts does not necessarily simplify the theory. The nature of singular concepts is not well understood, and this means that we cannot always count on them in constructing models of categorization. In fact, the problems of general and singular concepts are in some ways parallel (as Millikan, 2000, and Nozick, 1981, have pointed out): In the first case, we look for principles that unite discrete individuals into the same category; in the second, we look for principles that unite discrete glimpses or descriptions of an object into glimpses or descriptions of the same individual. Likewise, we use general concepts to keep track of properties that are constant or predictable from one instance to another, and we use singular concepts to keep track of properties that are constant from one phase of an object to another. These parallels again suggest that the representation of individuals may be more complex than might first appear.

In the first section of this article, we look at earlier cognitive theories of object identity. We argue that these theories are either not powerful enough to explain singular concepts or they rely on overly strong assumptions about the relation between singular and general concepts. We then outline a new model based on a notion of causal proximity that seems to handle some of the difficulties that previous theories face. As a test of the theory, we report studies in which participants have to decide whether one, both, or neither of a pair of possible successor objects is identical to an original object. We fit a quantitative version of the model to the results of two of these experiments and consider further predictions in other cases of choice between successors. Finally, we discuss issues that the theory raises about domain specificity and the transitivity of identity (Principle 1c), and we compare the model's advantages and disadvantages with those of earlier approaches.

### Theories of Singular Concepts

Before introducing our own theory of singular concepts, we outline three alternative ways of looking at this problem of identity of individuals across time. A first possibility makes use of the similarity between object descriptions. An alternative possibility, directed at the identity of concrete, physical objects, is to check on the spatial and temporal pathway that an individual follows. According to this proposal, people decide that an individual at an

earlier time is the same as one at a later time if and only if there is a continuous spatiotemporal path that connects them. Finally, people's notions of individual identity may depend on knowledge specific to the category to which it belongs. Perhaps people acquire criteria or rules for tracing identity as they learn what kind of thing an individual is. If so, then decisions about identity across time may be domain-specific—different for members of different basic-level categories.

We believe all three proposals have plausible elements and that people may actually use them in some settings to decide questions of identity. However, each has certain shortcomings that make it unlikely to serve as a general theory. In this section, we discuss their relative merits in turn, concentrating on theoretical strengths and weaknesses. We then revisit these proposals and consider their ability to predict new psychological data.

### *Similarity*

One possible answer to Question 1 is that individuals decide that items are identical by using knowledge of common and distinctive properties of the items to obtain an overall measure of the similarity between them. The items could then be judged as identical if the similarity exceeds some threshold or if the similarity between the first and second items is greater than the similarity between the first item and other possible contenders. We should take similarity seriously as a factor determining identity. First, it seems to influence perceptual impressions of identity. In apparent motion, for example, observers are more likely to see a single line segment moving than to see two distinct line segments if the segments are similar in sharing the same orientation (Ullman, 1979). Second, recognition of both individual words (e.g., Anisfeld & Knapp, 1968) and pictures (Bower & Glass, 1976) is sensitive to the similarity between the originally presented items and the test items.<sup>1</sup> In fact, it seems almost inevitable that similarity should play a role in judging the identity of objects. If a cat runs behind a couch and a very similar looking cat runs out the other side, we probably take this similarity as indicating a single cat in the absence of information to the contrary.

However, there are some general difficulties with a pure similarity theory, as may be clear from the fact that we sometimes count very similar items, such as identical twins or clones, as distinct individuals. First, properties of the items in question are likely to contribute unequally to judgments of identity. Aunt Florence's taste in music and other matters in 1970 (mostly Motown) may be vastly different from her taste in 2006 (mostly Mahler), so that her taste and preferences in 1970 and 2006 may differ in ways irrelevant to her identity. We, therefore, need a theory of which properties are relevant to judgments of identity, a variation of the question we started with. (Similarity-based theories of categorization have encountered the same sort of criticism; see Sloman & Rips, 1998, for a review of the status of similarity in cognitive models.) Second, similarity may presuppose identity, as Fodor and Lepore (1992) contend. If we use properties of Florence<sub>1970</sub> and Florence<sub>2006</sub> to establish the similarity between them, then we presumably must have some way to determine that these properties (or their instantiations) are the same. For example, if nearsightedness is one such property, we need to know that Florence's nearsightedness<sub>1970</sub> = Florence's nearsightedness<sub>2006</sub>. However, this shifts the problem from sameness of objects to sameness of

properties. Although this latter problem may be easier to solve than the former, there are no guarantees. Third, things change. We should expect some of Aunt Florence's properties to change in predictable ways over time, and although these changes make for dissimilarities, they should count for, rather than against, the possibility that a later stage belongs to the same individual as her earlier stages (Rosengren, Gelman, Kalish, & McCormick, 1991; Sternberg, 1982). An individual at age 3 would typically be shorter than, not the same size as, the same individual at 23. If individual  $x_0$  is 3'5" in 1986 and  $x_1$  is 3'5" in 2006, that would be evidence they were not identical. For these reasons, it is unclear how much similarity can help in specifying a theory of object identity. We later present some empirical evidence that bears on this issue (see Experiment 2).

### *Spatiotemporal Continuity*

According to the continuity view, we judge two individuals to be identical if we know that these individuals fall on the same unbroken spatiotemporal path. Florence<sub>1970</sub> = Florence<sub>2006</sub>, for example, if we can show that there is a continuous path or history linking the first to the second. This theory is similar to one that is sometimes offered for perceptual tracking in infancy. In one well-known experiment, 4-month-old infants repeatedly observed objects that moved back and forth on a stage that contained two opaque screens (Spelke, Kestenbaum, Simons, & Wein, 1995). In the continuous condition, the object passed behind the screen at the left, through an opening between the screens, and behind and beyond the second screen on the right. In a second, discontinuous condition, infants saw a similar motion pattern, but the object never appeared in the gap between the screens. In subsequent test trials, no screens were present, and the infants saw either one or two objects moving on the stage. Infants in the continuous condition tended to look longer during test at the two-object scene than at the one-object scene, whereas those in the discontinuous condition showed the opposite preference. This finding suggests that the infants interpreted the spatiotemporally continuous path through the gap as indicating the presence of a single object (and

<sup>1</sup> We note, however, that the relation between object identity and traditional recognition memory may not be straightforward. The standard recognition task is in some ways more about categorization than about object identity. If you are presented with the word *eggplant* and have to say whether it was on an earlier list, the correct answer is "yes" even if the word now appears in a different font, color, or modality. The correct answer depends on whether the original word and the current word are tokens of the same type, but, as we have already indicated, identity judgments are decisions about whether two appearances belong to the same token (i.e., are numerically identical). The relationship between perceptual object recognition and judgments of identity is potentially much closer. However, even here much of the research on object recognition is devoted to how people recognize objects as members of categories (e.g., horses) rather than on how they identify individuals (see Peterson, 2001, for a review of theories of object recognition). For example, the announced goal of Biederman's (1987) recognition-by-components theory is "to account for the initial categorization of isolated objects. Often, but not always, this categorization will be at a basic level, for example, when we know that a given object is a typewriter, a banana, or a giraffe" (p. 116). This is not to say that recognition is irrelevant to judgments of object identity but only that the relationships need to be carefully worked out.

were surprised when there were two) and interpreted the discontinuous path as indicating two objects (and were surprised when there was only one). (However, see Kuhlmeier, Bloom, & Wynn, 2004, for evidence that infants do not make similar inferences for humans moving in the same fashion.) There is also evidence (Stone, 1998) that the particular spatiotemporal path that an object takes can influence later recognition of that object; recognition is better if observers see the same path at test than if they see an equally informative alternative path. Clearly, psychologists have taken spatiotemporal continuity as critical for judgments of object identity.

Although the spatiotemporal continuity hypothesis is more substantial than the similarity proposal, there are counterexamples that suggest it does not always capture object identity. Armstrong (1980), Nozick (1981, pp. 655–656), and Shoemaker (1979) provide a thought experiment of this sort:

Imagine two machines: one capable of vaporizing an object and the other capable of materializing an object in an arbitrarily brief interval. Suppose, too, that these machines operate on completely independent schedules so there is no connection between one machine and the other. Then it is possible to conceive the first machine vaporizing a specific object (e.g., a chair) and the second machine, by chance, immediately materializing a qualitatively similar but distinct object without a temporal gap and in exactly the same spatial location. Under these circumstances, an observer would notice no change whatever, because nothing about their spatial or temporal position or their qualitative properties would distinguish the vaporized and materialized chairs from a single chair. However, in the imagined scenario, although there is an unbroken spatiotemporal sequence of chair stages, there are two chairs in play rather than one.

This example (which we will call *dual-ing machines*) suggests that people may always be willing to override purely perceptual information if they know enough about the facts of the case. For any imagined perceptual evidence pointing to one object, it may be possible to conceive Armstrong-Nozick-Shoemaker machines that substitute multiple objects.

One reaction to this example is that the machines do not truly preserve spatiotemporal continuity; there must be some break between the two chairs because they have different material composition. However, although there is a difference between the chairs, as we are about to discuss, there need not be any spatial or temporal discontinuity. It seems possible to envision the materializing machine outputting the second chair within any temporal interval  $\epsilon$  ( $\epsilon > 0$ ) after the disappearance of the first chair, and if so, this meets the standard definition of continuity. Of course, this example depends on contemplating sci-fi devices that may never actually exist, but the fact that we can make sense of the example suggests that we do not conceive of spatiotemporal continuity as guaranteeing identity over time.

Hirsch (1982) provides a second type of counterexample to the idea that spatiotemporal continuity is sufficient for identity. As Hirsch points out, there are indefinitely many spatially and temporally continuous sequences that do not count as a single object. The north half of a stationary cat from 10 to 11 p.m. is one such nonobject.

To be sure, the Armstrong-Nozick-Shoemaker and Hirsch examples do not show that all forms of continuity are irrelevant for identity. Intuitively, the reason there are two chairs rather than one in the dual-ing machine scenario is because the vaporized chair is

not connected to the materialized one in the same way as the successive stages of a single chair. In particular, there is no causal link between the vaporized and the materialized chairs. This intuition leads directly to the theory of identity that we propose later. However, incorporating causal relations takes us a significant step beyond spatial and temporal continuity.

It also seems doubtful that spatiotemporal continuity is necessary for identity. We could disassemble a computer into its individual circuit components, store the resulting hundreds or thousands of parts in separate locations, and then reassemble the parts later in yet another location but in precisely the same configuration. Under these circumstances, the earlier intact computer would seem to be the same object as the later reconstructed one. However, there is no continuous spatiotemporal path that links the two halves of the computer's existence, and this implies that identity is possible over apparent gaps in space or time (as Hirsch, 1982, argues on the basis of a similar example).<sup>2</sup>

The computer example should make us cautious about requiring continuity as a criterion of identity, but the example may also hint at another basis for singular concepts. Computers, tables, chairs, cars, and many other artifacts can survive complete disassembly and reassembly, but cats, robins, roses, and many other living things usually cannot survive total dismemberment, at least under currently available techniques. Some evidence that older children and adults recognize such a distinction comes from Hall (1998). Perhaps, then, identity over time is relative to the category to which an object belongs. It is a common theme in the categorization literature that knowledge of an object's category can provide theoretical information about the object, information that fuels inference and prediction (e.g., Medin, 1989). The theory we are about to take up extends this idea by supposing that category-level concepts also supply criteria for identifying category members from one moment to the next.

### Sortals

Certain concepts may determine rules for individuating and identifying their category members. The concept of tables, for example, may consist in part of rules for differentiating individual tables in a mass of tables and other objects and identifying each table over time. Of course, not all theories of concepts assume that concepts contain rules. However, some accounts, such as schema theories (e.g., Rumelhart & Ortony, 1977) or theory theories (e.g., Carey, 1985), are compatible with rules, and we are concerned

<sup>2</sup> It is possible to debate whether the computer exists during the time at which it is disassembled. Whether people view a disassembled object as the same individual may depend on the extent of the transformation (e.g., the number of resulting pieces or the size of these pieces). For instance, people may be more likely to believe that a scattered collection consisting of the disassembled top and legs of a table is still the same individual than a scattered collection consisting of the zillions of disassembled circuit components of a computer (see Gutheil, Bloom, Valderama, & Freedman, 2004, for relevant evidence). If the computer does not exist when its components are disassembled (as seems likely), then the example shows that objects can survive gaps in time. However, even if the computer continues to exist during its disassembled phase, it clearly does not exist as a spatially continuous entity. Therefore, transformations can preserve identity across (at least) spatial discontinuity.

here exclusively with this type of hypothesis (for reviews, see Murphy, 2002; Medin & Rips, 2005).

Philosophical work discusses this idea under the heading of sortals (Strawson, 1959, p. 168). A sortal is a count noun, like *table*, that is capable of singling out individual tables in a way that allows us to enumerate them. By contrast, an adjective like *black* denotes a property that does not by itself aid in separating and enumerating objects. We cannot count the black stuff that composes a black table, for example, because the total is indeterminate: It might be one (the table), five (the legs + the top), six (the legs + the top + the table), or more. Nouns like *table*, *leg*, and *top*, however, do provide the resources we need to get a determinate answer to the question of how many items of this type are on hand. Orthodox sortal theories assert that there are no individuals at all, apart from the sortal concepts that carve them out and establish their beginnings and endings. As Dummett (1973, p. 179) put 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.” In what follows, we use the term *sortal* for linguistic expressions (i.e., for certain count nouns), *sortal concept* to refer to the associated mental representation, and *sortal category* to refer to the referent of the sortal.

It is a natural next step to enlist sortals in tracing the history of individuals. An individual object, such as a cat or a table, can undergo a variety of changes in its properties, whereas other property differences are not compatible with identity. This distinction between possible and impossible differences for an individual then determines, at least in part, the identity of that individual. An individual  $x_0$  cannot be identical to an individual  $x_1$  if  $x_0$  and  $x_1$  exhibit a change that is not compatible with objects of  $x_0$ 's type. Which changes are possible and which are impossible varies across types of objects. Some changes such as total disassembly and reassembly may be possible for a table but not for a cat. So the issue of whether  $x_0$  continues as  $x_1$  (and whether  $x_0 = x_1$ ) depends on what  $x_0$  and  $x_1$  fundamentally are. According to Wiggins (2001), sortals (e.g., *table* or *cat*) provide the answer to this “What is it?” question. Thus, the meaning of a sortal supplies information not only about how to distinguish the things it applies to but also about the identity of these things over time.

One advantage of the sortal theory is that it handles some issues that are problematic for continuity theories. Consider, for example, a car that loses a hubcap on a bumpy road. Although both the hubcapless car and the carless hubcap are continuous with the original item, the sortal *car* applies to the initial object and dictates that it is the hubcapless car that is identical to the original car. Advantages like this one suggest that it might be useful to incorporate the sortalist insights in psychological explanations of object identity, so we need to examine carefully attempts of this kind.

*Sortalist approaches in psychological theories.* For researchers who see deficiencies in pure similarity and continuity accounts, sortals help fill an explanatory gap by providing a source of rules that people can use to keep tabs on things. As Carey (1995, p. 108) puts it, “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. . . Next consider whether a given entity is the same one as

we saw before. Again, we must be supplied a sortal to trace identity.” Although we know of no detailed psychological model of how people represent sortal categories' rules for distinguishing and tracking objects, the outlines of the approach are reasonably clear. The sortal concept must specify which properties of an individual category member can change over time (and in what way) and which properties are fixed. In this respect, the sortal theory is similar to earlier proposals about categories in which a concept includes information about the category's variability (e.g., Fried & Holyoak, 1984; Walker, 1975). However, in this case, what is specified is the allowable variability within individual category members rather than between the members. Thus, a sortal concept can provide at least a partial answer to Question 1, which we can formulate as follows:

Principle 3. Object  $x_0$  in situation  $S$  is identical to an object  $x_i$  in  $S'$  only if  $x_0$  and  $x_i$  exhibit changes that are compatible with  $x_0$ 's sortal.

In this vein, Macnamara (1986) assumed on theoretical grounds that when children learn a proper name for an object, they interpret the name with the help of the object's sortal concept. The sortal concept, which Macnamara took to be a prototype or perceptual gestalt of a category, provides criteria for individuation and identity that support correct use of the proper name. The same considerations apply to the use of personal pronouns, such as *I* and *you* (Oshima-Takane, 1999).

An example of how sortals play a role in explaining identity over time comes from Xu and Carey's (1996; Xu, Carey, & Quint, 2004) experiments on object individuation. In these studies, infants viewed a screen from which objects emerged, either at the right or left. On a particular trial, an infant might see, for example, a ball emerge from the right side of the screen and then return behind the screen. A short time later, a cup emerges from the left side of the screen and returns behind the screen. This performance is repeated a number of times with the same two objects. The screen is then removed, revealing either a single object (e.g., the cup) or two objects (cup and ball). These experiments show that when the screen is removed, 10-month-old infants look no longer at the scene with one object than at the scene with two (relative to baseline performance). By contrast, 12-month-olds look longer at the one-object tableau as long as the objects are from different basic-level categories. (If the objects are from the same category, e.g., two cups, but have contrasting properties, e.g., different shapes, even 12-month-olds fail to look longer at the one-object scenes; see Xu et al., 2004.) Xu and Carey interpret this to mean that the younger infants do not expect to see two objects and so are no more surprised by one than by two in this context. Older infants, however, can use their knowledge of sortal concepts to make the discrimination.

Several factors allow younger infants to anticipate two objects correctly. First, if the 10-month-olds are able to inspect simultaneously both the cup and the ball before the start of the trial, then they do stare longer at the one-object scene. Second, if the experimenter labels the two objects differently (“look a blicket” vs. “look a gax”) while they are moving back and forth, 10-month-olds again perform correctly (Xu, 2002). This combination of results suggests, according to Xu and Carey (1996), that younger infants can use spatial or verbal cues to individuate the objects; without these cues, they are unable to anticipate the presence of

two objects, because they do not know that balls do not morph into cups while briefly out of sight.

According to Xu and Carey (1996), the younger infants who fail the “Is-it-one-or-two?” task lack knowledge of sortal concepts (e.g., *cup*, *ball*) that would allow them to individuate the objects conceptually. Because this individuating information is supposed to be a crucial part of the meaning of sortals, these infants do not know these meanings; they do not have adultlike concepts for even basic-level categories such as cups.<sup>3</sup>

*Evidence concerning sortals.* Carey and Xu (2001; Xu, 2003, 2005) maintain that infants acquire the meaning of sortals, such as *cup* and *ball*, at about 12 months of age and that the sortals are responsible for older infants’ and adults’ correct performance in the “Is-it-one-or-two?” task.<sup>4</sup> We, therefore, need to examine whether *cup*, *ball*, and similar count nouns play this identifying role.

One way to investigate this issue takes advantage of Wiggins’s (1997, 2001) contention that the sortal for a particular object is the term that answers the question, “What is it?” Brown (1958) and Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) have claimed that words for basic-level categories, such as *cup* or *ball*, usually provide the answer to this question; so it may be possible to check whether knowledge of basic-level categories gives people the means to identify objects. In one attempt of this kind, Liittschwager (1995) gave 4-year-old children illustrated stories about people who were magically transformed to different states. The transformations ranged, across trials, from simple within-category changes in properties (e.g., from a clean to a dirty child) to more extreme cross-category changes (e.g., from a girl to a cat or from a woman to rain). For each type of transformation, participants were to decide whether the transformed object could still be called by the name of the original person (e.g., “Do you think that now this is Ali?”). According to sortal-based theories, objects cannot maintain their identity across changes in sortal categories; so participants should use the same proper name only if the transformation is within the basic-level category person. The results of this study showed that as the transformational distance increased between the original person and the final product, participants were less willing to apply the proper name. However, there was no discernible elbow in this function at the sortal category boundary: the boundary between persons and nonpersons. According to Liittschwager (1995, pp. 33–34), the data “provide little support for Macnamara’s (1986) position that proper names should be maintained across changes up to (but not beyond) the basic level.”

Blok, Newman, and Rips (2005) report a related finding in an experiment that also uses transformation scenarios. Participants (college students) read stories about an individual (e.g., Jim) who has a severe traffic accident in 2020 and who must undergo radical surgery. In the condition most relevant for present purposes, participants learned that Jim’s brain was transplanted to a different body. On some trials, scientists placed the brain in “a highly sophisticated cybernetic body,” whereas on others they placed it in a human body that scientists had grown for just such emergencies. In each case, Jim’s old body was destroyed. The stories described the operation as successful in allowing the brain to control the new body, but participants also learned that Jim’s memories either survived the operation intact or did not survive. After reading the scenario, using a scale ranging from 0 to 9, participants rated their agreement with each of two statements: (a) The transplant recip-

<sup>3</sup> One possible issue, and a source of conflict with sortal theories in philosophy (e.g., Wiggins, 2001), is that sortals like *cup* or *ball* should also be necessary to individuate objects that appear together in the perceptual field. The evidence from Xu and Carey’s experiments (Xu & Carey, 1996; Xu et al., 2004), however, is that younger infants do perform correctly when they have the advantage of previewing the objects. To explain this difference in performance, Xu and Carey argue that even the younger infants have a high-level sortal concept, equivalent to the concept “physical object,” which Spelke (e.g., 1990; Spelke, Gutheil, & Van de Walle, 1995) has posited to explain infants’ object tracking. This concept provides the sortal information that infants use in the preview condition (i.e., The same object can’t be in two different places simultaneously.) As Xu (1997, p. 369) states, “for both adults and young infants, there is nonetheless a sortal *physical object*, which is more general than *person*, *car*, or *tree*. A physical object is defined as any three-dimensional, bounded entity that moves on a spatiotemporally continuous path” (see also Carey, 1995; Carey & Xu, 1999). However, sortal theories in philosophy typically hold that terms like *thing*, *object*, *physical object*, *space occupier*, *entity*, and so on, are not sortals, despite their count noun syntax, because they do not provide identity conditions (e.g., Hirsch, 1982, p. 38; Wiggins, 1980, p. 63; Wiggins, 1997, p. 418). Just as we cannot count the black stuff that constitutes a black table, we cannot count the physical objects that constitute it; the number could again be one (the table), five (the legs and top), six (the legs, top, and the table), and so on. One way to square sortals with Spelke’s physical objects is to note that Spelke’s object concept is more specific than the ordinary notion of a physical object. Many things that we single out as objects do not move independently and are not spatially separated from their backgrounds (as Hirsch, 1997, and Wiggins, 1997, have pointed out). Trees, mountains, houses, fences, fire hydrants, and sidewalks, among other things, are typically fixed in place and would fail to trigger an object concept that is sensitive only to movement and spatial isolation. Similarly, nonmoving parts of larger wholes often qualify as objects in the everyday sense but not in the sense of independently moving, spatially separated entities. We speak of legs of tables, fenders of cars, handles of mugs, organs of animals, and other parts as objects in their own right, despite the fact that they usually occupy a fixed position with respect to the relevant larger entity. A Spelke-type object concept cannot pick out such objects. For this reason, it seems best to regard this concept as corresponding to a kind of primitive or proto-object (sometimes called a Spelke object). Could *proto-object* be a sortal? Because the parts of a table, for example, are not proto-objects (they do not move on their own), counting the proto-objects that constitute a table does not pose the problem that counting physical objects does (Carey & Xu, 1999; Xu, 1997). A table is a single proto-object. (For arguments against the idea that *proto-object* is a sortal, see Ayers, 1997; Hirsch, 1997; Wiggins, 1997.) The idea that both proto-object and lower level terms like *cup* simultaneously function as sortals still conflicts with strong sortal theories (e.g., Wiggins, 2001) in which only a single sortal captures all the identity conditions for a particular object. However, this idea is compatible with the weaker Principle 3, in which a sortal simply contributes some conditions on an object’s identity. For the moment, then, we will take Principle 3 as the official (psychological) version of the sortal theory and see what additional evidence we can bring to bear on it. We also consider theories weaker than Principle 3 later.

<sup>4</sup> Experiments following Xu and Carey (1996) have found cases in which infants 10 months of age or younger are able to perform correctly in simplified versions of the “Is-it-one-or-two?” task (e.g., Wilcox & Bailargeon, 1998; Xu & Baker, 2005) or versions that center on humans versus other objects (Bonatti et al., 2002). The exact age at which infants succeed at such tasks is not of central interest here; however, some of the explanations for this early success do bear on the question of what knowledge they draw on when they anticipate two objects versus one object. Carey and Xu (2001, p. 194) argue that “when spatiotemporal evidence

ient is Jim after the operation, and (b) the transplant recipient is a person after the operation.<sup>5</sup>

The results from Blok et al. (2005) show a dissociation between identity and category judgments. Figure 1 displays the mean agreement ratings as a function of whether the story described the brain transplanted to a robot or to a human body and also whether the memories survived or did not survive the operation. Participants were more likely to agree that the postoperative recipient was still Jim if Jim's memories were preserved. However, there was a much smaller effect of whether these memories were embodied in a human body or in a robot body. Agreement about whether the end product was a person, however, showed the opposite pattern. Participants were more likely to think the transformed object was a person if it had a human body rather than a robot body, but they relied less heavily on whether Jim's memories remained intact. This combination of effects produced the finding that when Jim's memories survive in a robotic body, participants are much more likely to think that the transformed individual is Jim than that it (!) is a person.

Psychological versions of sortal theories seem at odds with this outcome. These theories subscribe to Principle 3, because, for example, this is the principle that they use to explain 12-month-olds' correct performance in the "Is-it-one-or-two?" task. Principle 3 is logically equivalent to the statement that objects differ if they exhibit a change incompatible with the original object's sortal. However, when Jim's brain is transplanted to a robot body with his memories intact, participants judge him to have undergone a change incompatible with his sortal (he is no longer a person) while remaining Jim. This contradicts Principle 3 and casts doubt on sortal theories. Of course, the category of persons may be special and perhaps unrepresentative of other kinds of objects with respect to identity (e.g., Bonatti, Frot, Zangl, & Mehler, 2002; Kuhlmeier et al., 2004; Sternberg, Chawarski, & Allbritton, 1998, Experiment 3). However, similar dissociations appear in experiments using other categories (Blok et al., 2005, Experiment 2). A new study of this type appears later as Experiment 1.

### A Causal Continuer Theory of Object Identity

We consider now a model of identity judgments that draws on some of the elements of the earlier views we have just surveyed

---

does not favor one solution over another, infants can use featural differences for object individuation" (see also Xu, 2003, 2005). These featural or property differences may be responsible for correct performance of the younger infants in Wilcox and Baillargeon (1998) and Xu and Baker (2005). However, in Xu and Carey's original (1996) task, spatiotemporal information from the moving objects (i.e., trajectory information suggesting a single object) overrides featural differences that would otherwise serve to distinguish the objects, causing errors for the younger infants. Older infants are able to marshal sortals that, in turn, overcome the misleading spatiotemporal facts. However, featural differences (e.g., color and size changes) are precisely the kinds of properties that do not individuate objects, according to the philosophical theories of sortals described earlier (e.g., Strawson, 1959; Wiggins, 2001). Recall our earlier discussion of *black* versus *table*. To the extent that infants can use properties (without the support of underlying sortals) to distinguish the items in these experiments, the very difference between sortal and nonsortal predicates is placed in doubt.

but combines them in a new way. The model attempts to answer Question 1 by describing the cognitive processes people go through when they have to decide whether an individual object,  $x_0$ , existing at one time is identical to one of a set of candidate objects,  $x_1, x_2, \dots, x_n$ , existing at a later time. The model derives from Nozick's (1981) closest continuer theory, but we recast the proposal here as a descriptive psychological account. We intend the model to help explain how people judge object identity, and we do not take a stand on whether it applies correctly to possible nonpsychological (metaphysical) aspects of identity. We first describe the theory in outline and then apply it to some new experimental data that we have collected to test a quantitative version of this approach. In the following sections, we discuss extensions of the theory and examine potential limitations.

### The Causal Continuer Theory

As its name implies, Nozick's (1981) closest continuer theory commits itself to the idea that the object identical to the original  $x_0$  is the one that is, in some sense, closest. The later manifestation of  $x_0$  is the one most like  $x_0$ , according to some weighting of its properties. However, unlike the similarity approach, which we discussed previously, the present theory determines closeness within a framework of causal principles. To emphasize these restrictions as well as some of our own modifications, we refer to the model proposed here as the *causal continuer* theory.

Causality is important in this context because the theory's chief idea is that the continuer of the original object must be a causal outgrowth of that original. We illustrate this idea in the following story of the missing chair.

Suppose you own a favorite chair with a particular color and shape. One day you regrettably leave the chair in one of the classrooms in the department. When you return to reclaim it, it is nowhere to be found. The following week you spot two different chairs that look just like yours. One is sitting in the office of Professor A and the other in the office of Professor B. Which, if either, of these chairs is yours? Similarity is clearly unable to decide the case. Spatiotemporal continuity might be helpful if you could establish that there is a continuous spatial pathway from the chair in the classroom to the chair in one of the offices. However, suppose that on investigating you find that Professor A, who had never seen your chair, happened to construct one of the same shape and color. Professor B, however, has disassembled the chair he found in the classroom, stealthily moved the parts to his office one at a time, and reassembled the chair. In neither case is there a spatially continuous path from your chair to the resulting chairs. In particular, while Professor B was transferring the chair parts to his office, the parts were disconnected and spatially scattered. Nevertheless, there is a clear intuition that it is the chair in Professor B's office, and not the one in Professor A's, that is yours. An obvious causal relation links each step in the transition from the

---

<sup>5</sup> We assume, along with Liittschwager (1995) and others, that proper names like *Jim* are rigid designators that always refer to the same individual across situations or possible worlds; see Kripke (1972). Participants who state that the transplant recipient is no longer Jim are, therefore, affirming that the recipient is no longer the same individual.

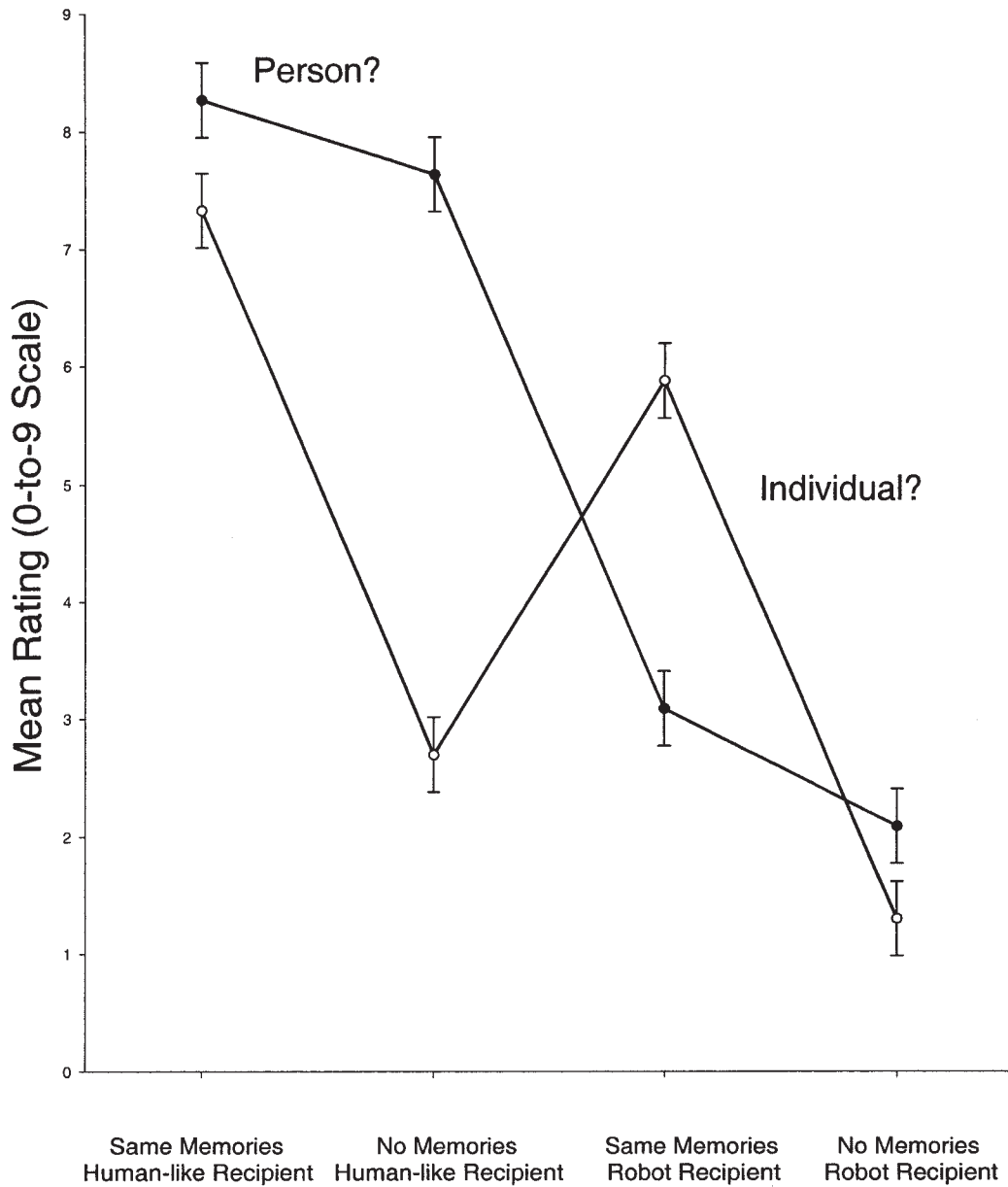


Figure 1. Mean agreement ratings (0-to-9 scale) for the statements that the transplant recipient is “still Jim” (open circles) and “still a person” (closed circles). The x-axis represents four versions of the accident story. Error bars indicate  $\pm 1$  standard error of the mean (from Blok et al., 2005).

chair in the classroom to the chair in Professor B’s office, but there is no such causal relation between your chair and Professor A’s.

The important role that causality plays in the theory goes along with the intuition that causal forces are central in producing an object, maintaining it through time, and eventually destroying it. In this respect, the causal continuer theory is akin to psychological essentialism (Gelman, 2003; Medin & Ortony, 1989), which also emphasizes the role of causality in people’s thinking about natural kinds. It also agrees with some versions of psychological essentialism in supposing that separate causal factors are responsible for

category membership and individual persistence (Gelman, 2003; Gutheil & Rosengren, 1996). However, the present theory takes no stand on the existence of a unique, distinctive cause that would answer to the notion of an essence. For present purposes, the existence of an object may be a function of many conspiring causes, some overlapping those of other objects (see Rips, 2001; Sloman & Malt, 2003; Strevens, 2000). Similarly, the theory makes contact with recent models of categories that emphasize the role of causality in category structure (e.g., Ahn, 1998; Rehder & Hastie, 2001). For reasons that we have mentioned earlier in



connection with sortals, however, we assume that the causes responsible for an individual's persistence may include not only those associated with its basic-level category but also the larger set of background causes that govern the individual's environment.

A second aspect of the theory is that, in determining a continuer, we cannot select something that is arbitrarily far from the original. In some later situations, there may be no object that qualifies as identical to the one with which we started. It may be, for example, that although there are objects at a later time that causally stem from the original, the causal connections are so attenuated that none of the objects can serve as a continuer, and the original object thereby goes out of existence. If a book is ripped apart into its covers and its individual pages (each page separated from the others), then each of the resulting pieces maintains a causal connection to the original, but the connection may not be great enough to qualify any of the pieces (or their sum) as the book. Similarly, the causal changes between the original object and a later one cannot be too abrupt. Although the dead remains of an animal causally stem from its living state, the transition is not smooth enough to allow the remains to serve as a continuer of the organism.

Finally, it may sometimes happen that there are two or more objects at a later time that are causally close enough to the original to be potentially identical. In that case, the causal continuer theory specifies that only the closest of these options is identical to the original. It is this last aspect of the theory that allows it to address some traditional puzzles about identity. The most famous of these is due to Hobbes (1655/1839–1845), which we can paraphrase in the following ship of Theseus scenario:

A wooden ship was repaired over a long interval by removing individual planks one at a time and replacing them at each step with new ones. This process continued until none of the old planks remained, and the ship consisted entirely of new planks. However, the old planks were stored and then reassembled exactly as before. Two ships exist at this later point, each of which could claim to be the original ship: the one with old planks and the one with new planks. Which, if either, is Theseus's ship?

The causal continuer model can afford to be neutral with respect to the choice between the ships (see Nozick, 1981). Both the resulting ships—call them Old Parts and New Parts—can be seen as causal outgrowths of the original. However, New Parts enjoys closer temporal continuity with the original, whereas Old Parts has greater overlap in material composition. Whether we deem Old Parts, New Parts, both, or neither as Theseus's ship will then depend on how we weigh these two factors. In this respect, the model resembles standard models in decision theory in which different people may have different weighting functions. The model does not make an a priori decision among the options, but it does explain the uncertainty we may feel about the choice. Both composition and temporal overlap are typically important and perfectly correlated in identity judgments about ordinary ships. Both are diagnostic of the causal forces that support a ship's existence. Hobbes's story unconfounds these factors, forcing us to consider them separately, and it is this new demand for independent weighting that creates the puzzle. In the same manner, the model also accounts for the intuition that either Old Parts or New

Parts would unambiguously be Theseus's ship if the other were out of the picture. For example, if the original ship were simply disassembled and reassembled, we probably would not hesitate to identify it with the ship of Theseus. Similarly, if the parts of the original were gradually replaced with no reassembly of the old parts, then the ship of Theseus would be the repaired ship. What creates indecision in Hobbes's problem is the competition between Old Parts and New Parts for being the closest or best option.

### *Some Implications of the Causal Continuer Approach*

We can think of the causal continuer theory as imposing a two-part decision process on judgments of identity. To determine which of a set of objects at a later time is identical to an original, we perform two steps. Step a: We consider only those later objects whose connection to the original exceeds some threshold (no other objects can be continuers). Step b: Within the range of close-enough objects, we select the closest as the one identical to the original.<sup>6</sup> It may seem natural to assume that people carry out Step a before Step b, but the opposite ordering is also possible. People may identify the closest object before determining whether that object is close enough to be identical. The mathematical model that we present later does not opt for one order over the other.

We also note that Step b makes the decision process context sensitive. An item that is closest in one situation may not be closest in another if the second situation contains an even closer object. Criticism of the closest continuer theory in philosophy has focused on this context sensitivity (e.g., Noonan, 1985; Williams, 1982). According to these criticisms, the question of whether  $x_0$  is identical to  $x_1$  cannot depend on the presence of an individual  $x_2$  that may also exist at the same time as  $x_1$ . The appeal of this idea (sometimes called the *only-x-and-y* principle) stems from the intuition that the identity of an individual is an intrinsic matter and, therefore, cannot be affected by the presence of other things. However, whether or not this is a correct metaphysical rule (Nozick, 1981, argues against it), considering alternatives seems an inevitable part of judging (or inferring) the identity of objects, which is the process in Question 1 that we hope to clarify. This context sensitivity is on a par with similar effects in judgments of similarity (e.g., Tversky, 1977) and choice (e.g., Shafir, Simonson,

<sup>6</sup> Nozick (1981) also requires that causal proximity holds in both directions. Not only must the later object be the closest continuer of the original, but the original must also be the closest predecessor of the later one (the objects must be mono-related, in Nozick's terminology). Violations of this condition can occur, for example, in a situation in which there are two objects,  $x_0$  and  $x_0'$ , at an earlier time and a single potential continuer,  $x_1$ , at a later time. Then  $x_1$  may be the closest continuer of  $x_0$ , but  $x_0'$  the closest predecessor of  $x_1$ . In that case  $x_0$  cannot be identical to  $x_1$ , according to the restriction in question. We take monorelatedness to be a reasonable restriction on identity. In the experimental data we report here, however, violations of the closest-predecessor relation do not arise, and we do not discuss it further. Another potential qualification may occur in the case of people and other sentient beings, in which psychological continuity may take precedence over physical causal continuity in determining identity (see Kolak & Martin, 1987). We do not know whether people think of psychological continuity as existing apart from an underlying causal-physical basis and, if so, whether they privilege the mental over the physical. However, if they do, the causal continuer model may have to be restricted to nonsentient individuals.

& Tversky, 1993). We take up empirical evidence for context sensitivity later in discussing the results of Experiment 2.<sup>7</sup>

In Nozick's (1981) theory, the closest continuer must be closest in an absolute sense; no ties are allowed. For example, if an amoeba divides in such a way that the two descendants are equally close to the common parent, then the parent cannot be identical to either descendant. The reason for this additional restriction is not difficult to understand. The two descendants,  $x_1$  and  $x_2$ , do not seem equal to each other, because each can go its own way, acquiring different properties after the division that produced it. However, if the parent  $x_0$  is then equal to both the descendants, the result is an intransitivity:  $x_1 = x_0$  and  $x_0 = x_2$ , but  $x_1 \neq x_2$ . However, similar apparent intransitivities arise in certain perceptual situations (Ullman, 1979), and for this reason we leave room for the possibility of ties in judgments about conceptual identity.<sup>8</sup> If such judgments do exist, we can then consider whether they are best seen as true intransitivities (perhaps a type of performance error) or reflect a different interpretation of the objects (the values of  $x_0$ ,  $x_1$ , and  $x_2$ ) that can salvage transitivity.<sup>9</sup>

### An Examination of the Causal Continuer Theory

In testing the theory, we have two goals. The first is to show that causal factors rather than similarity or sortal category membership dominate judgments of object identity when these factors are contrasted. Experiment 1 compares causal predictions with those of the sortal theory, and Experiment 2 compares causality and similarity. Our second goal is to test a quantitative version of the causal continuer approach. As we have just noticed, in this model, identity judgments involve a double comparison process in the general case: The identical object must be causally close enough to be the original and must be closer than other close enough alternatives. To test the theory, we need a situation in which at least two contenders are available, as in the Ship of Theseus problem (see Hall, 1998, for other experimental analogues of this scenario). In both Experiments 1 and 2, we include conditions of this sort and use the model to predict exact response distributions for these conditions. We later attempt to establish the generality of the model by extending it from the temporal to the spatial domain, and we consider problems that arise in our modeling endeavor.

#### Experiment 1: Individual Persistence Across Transformations

To find out how well the causal continuer theory handles people's identity judgments, we need an experimental task that gives participants a choice between potential continuers and varies the causal distance between the continuers and the original object. Because we are also interested in the effects of category membership (as an additional test of the sortal theory), two or more categories must be involved. These requirements are difficult to satisfy with everyday objects, but we can approximate them in stories about hypothetical transformations, as in earlier research on concepts and categories (e.g., Blok et al., 2005; Gelman & Wellman, 1991; Johnson, 1990; Keil, 1989; Liittschwager, 1995; Rips, 1989). We report experiments with more naturalistic scenarios later in this article.

The stories we used in this experiment are similar to those in some philosophical discussions of identity (e.g., Lewis, 1983;

Nozick, 1981; Parfit, 1984; Perry, 1972) and described a machine that could copy and transfer objects from place to place on a particle-by-particle basis. The copied particles are transmitted to a new location and put back together according to a blueprint of the original. The particles of the original are entirely destroyed in the copying process. Thus, there was no spatiotemporal or material continuity between the original and the copy, but the copy causally stems from the original by means of the duplicating process.<sup>10</sup> (This explicit causal relation distinguishes this setup from the dual-ing machine example described earlier.) Each trial of the experiment described a different hypothetical transformation of this sort, and participants' task was to make two decisions about the resulting copies: (a) whether the copy is the same object as the original and (b) whether the copy is in the same category as the original.

To separate causal from sortal influences, we varied two aspects of the transformation. One factor was the proportion of particles in the copy that the original object causally produced (through the duplicating process). Across trials, we told participants that this proportion took the value 0%, 25%, 50%, 75%, or 100%. The other variable was the source of the remaining particles when less than 100% stemmed from the original. On half of these trials, the residual particles were said to come from another member of the original's basic-level category; on the rest of the trials, the residual

<sup>7</sup> It is an interesting further question, however, whether people would endorse the context-sensitive principle or the only- $x$ -and- $y$  principle if they were probed directly about the nature of identity. It is quite possible that people's identity judgments could be context sensitive, even though the same individuals believe that identity does not depend on the presence of other potential continuers. In this article, we focus on the identity judgments themselves and postpone to further research the question of people's (metacognitive) theories of identity.

<sup>8</sup> In the simplest situation of this type, an observer sees a central dot,  $x_0$ , in an initial display. This dot disappears, and then two dots,  $x_1$  and  $x_2$ , appear in a second display, with  $x_1$  and  $x_2$  located on either side of, and equally distant from, the position  $x_0$  had occupied. If the interstimulus interval is appropriate and the observer fixates  $x_0$ , then he or she sees simultaneous movement toward both  $x_1$  and  $x_2$  (Ullman, 1979). However, on the assumption that motion correspondence implies identity (Kahneman, Treisman, & Gibbs, 1992), we get a potential violation of the transitivity relation in Principle 1c. If  $x_1 = x_0$  and  $x_0 = x_2$ , then transitivity yields  $x_1 = x_2$ . However, it appears that  $x_1 \neq x_2$  because these two dots are in separate locations in the second display.

<sup>9</sup> Although Nozick's (1981) model blocks intransitivities of the sort just described, there are other ways in which both Nozick's model and our own allow intransitive judgments. Suppose object  $x_0$  exists at time  $t_0$ ,  $x_1$  at  $t_1$ , and  $x_2$  and  $x'_2$  at  $t_2$ . Then  $x_1$  might be the closest continuer of  $x_0$ , and  $x_2$  the closest continuer of  $x_1$ , but  $x'_2$  might be the closest continuer of  $x_0$ . In the context of psychological judgments, intransitivities may also arise if an observer cannot distinguish  $x_0$  from  $x_1$  or  $x_1$  from  $x_2$  (e.g., because of limits on acuity) and, therefore, decides  $x_0 = x_1$  and  $x_1 = x_2$  but can distinguish  $x_0$  from  $x_2$ . In the experiments reported here, however, we consider only situations involving two time points and alterations that are readily distinguishable.

<sup>10</sup> The machine destroys the spatial continuity of the object because the particles constituting it are not spatially adjacent during the transmission process. Does it also destroy temporal continuity? This depends on whether the object can be said to exist during transmission. Our intuition is that this is not the case, for the reasons mentioned in footnote 2, so there is a temporal gap in the object's existence if it survives the transformation.

particles came from a member of a different basic-level category. We predicted that the source of these residual particles would affect participants' judgments about whether the copy was still a member of the initial basic-level category but might have only minimal effects on identity judgments. The proportion of particles copied from the original, however, should influence identity judgments but be unrelated to category judgments. This dissociative pattern would echo that of Figure 1 and would imply that identity decisions need not follow category (sortal) decisions. The sortal theory implies, however, that factors that cast doubt on the object's basic-level category should also cast doubt on its identity (see Principle 3). In particular, if participants judge that the source of the residual particles changes the object's basic-level category, the same factor should affect judgments of identity.

### Method

*Stimulus items.* In a first block of trials, the instructions told participants that the machine had made a single copy of the particles, and the participants decided whether that copy was identical to the original and whether it was in the same category as the original. In a second block of trials, the instructions stated that the machine constructed two copies. Participants then decided whether one, both, or neither of these copies was identical to the original and whether one, both, or neither was in the original's category. As we have noted, in the one-copy condition, the copy could contain 0%, 25%, 50%, 75%, or 100% of the particles copied from the original. In the two-copy condition, each copy could independently contain any of the five percentages just mentioned, with the residual particles again coming from a different object. For example, participants might learn that one copy included 50% particles coming from the original object and 50% from a separate object, whereas the second copy included 75% particles from the original and 25% from the separate object. (The percentage of particles from the original need not add to 100%, because the machine was said to have made two complete batches of particles.) In the context of this experiment, the percentage of particles from the original object provides a measure of the causal distance between a copy and the original. The instructions specified that the particles of the original object were destroyed to ensure that there was no material continuity between the original and the copy. No mention was made of whether the particles from the separate object were originals or copies.

In each story, the original item was a lion (called "Fred"), and the residual particles were either from a second lion ("Calvin") or from a tiger ("Joe"). Thus, in the one-copy condition, participants might learn on one trial that the newly constructed creature contained 75% particles copied from Fred and the remaining 25% from the same-category member, Calvin. On another trial, the creature contained 75% particles copied from Fred and the remaining 25% from the different-category member, Joe. In the two-copy condition, both copies had residual particles from the second lion or both had residual particles from the tiger. The stories gave different labels to the alternative basic-level categories (lion vs. tiger) and also to the different individuals contributing particles to the copy (Fred vs. Calvin/Joe). Thus, any influence of an alternative label should be the same for decisions about whether the copy was the same individual (still Fred?) as for decisions about whether the copy was in the same basic-level category (still a lion?). This speaks to a possible methodological issue in the studies by Blok et al. (2005) and Liittschwager (1995), in which there were no alternative proper names.

*Procedure.* The instructions told participants to imagine that there was a machine consisting of a disassembler, a computer, and an assembler: "An object is placed in the disassembler and information is gathered about the kinds of particles that make up the object and their location in space. As the disassembler reads information about the object, each particle is destroyed. This information is then fed into a sophisticated computer, which makes an

exact copy of each molecule. The assembler then creates an object by arranging the stored particles in the way specified by the computer." In the one-copy condition, participants received nine scenarios that differed in the percentage of particles coming from the original object and in the source of the residual particles. (There were 9 rather than 10 scenarios, because when 100% of particles were from the original, there were no residual particles and thus no possible difference in their source.) On each trial, participants made separate decisions about whether the outcome of the transformation was the same individual (they chose between "is Fred" or "is not Fred") and whether it was a member of the same category ("is a lion" or "is not a lion"). In the two-copy condition, participants received the 30 trials. For each story, they again made an individual decision (they selected one of the following: "Only Copy A is Fred," "Only Copy B is Fred," "Both copies are Fred," or "Neither copy is Fred") and a category decision ("Only Copy A is a lion," "Only Copy B is a lion," "Both copies are lions," or "Neither copy is a lion").

A computer presented the instructions and scenarios on a monitor, and participants registered their responses by pressing designated keys on a keyboard in front of them. Approximately half the participants made the individual decisions first on all trials, and the remainder made the category decisions first. Within each of the one-copy and two-copy blocks, the scenarios appeared in a new random order for each participant. We tested the participants one at a time. They proceeded through the trials at their own pace, advancing to the next screen by pressing the space bar and taking about 45 min to complete the session.

*Participants.* The participants were 22 Northwestern University students who took part in order to fulfill a course requirement in introductory psychology. Because of an error, however, the one-copy data from 1 participant were lost, leaving a total of 21 participants in that condition.

### Results

*One-copy condition.* The one-copy data provide evidence about which of the experimental factors affect decisions about individual identity and about category membership. The results appear in Figure 2, and they exhibit a dissociation confirming the results from Blok et al. (2005) that we discussed earlier. Figure 2 shows that the larger the percentage of particles from the original individual, the more likely participants are to say that the copy is the same as the original. Of the 21 participants in this condition, 19 made more "same" decisions when the percentage of original molecules was high (75% or 100%) than when it was low (0% or 25%); the 2 remaining participants produced tied scores. In the 0–75% range, the slope in Figure 2 is fairly gradual but still amounts to an increase of 19 percentage points. There is no effect, however, of whether the residual particles are from a member of the same category or of a different category.

By contrast, judgments of whether the copy is in the same category as the original have a very different shape. When the residual particles are from a member of the same category, participants agree that the copy is also a member of that category on 89% of trials. When some of the particles are from a member of a different category, however, agreement falls abruptly to 33% and does not vary with the proportion of particles from that category member. Sixteen participants showed this effect; the remaining 5 gave tied scores. Because we are dealing with binary choices, we used a repeated measures analysis of categorical data to confirm these effects, and we report the Wald statistic,  $Q_w$  (Koch, Landis, Freeman, Freeman, & Lehen, 1977). To reduce the number of parameters the analysis must estimate, we included only those trials in which 0% or 75% of particles were from the original individual. This analysis confirms that both the percentage of

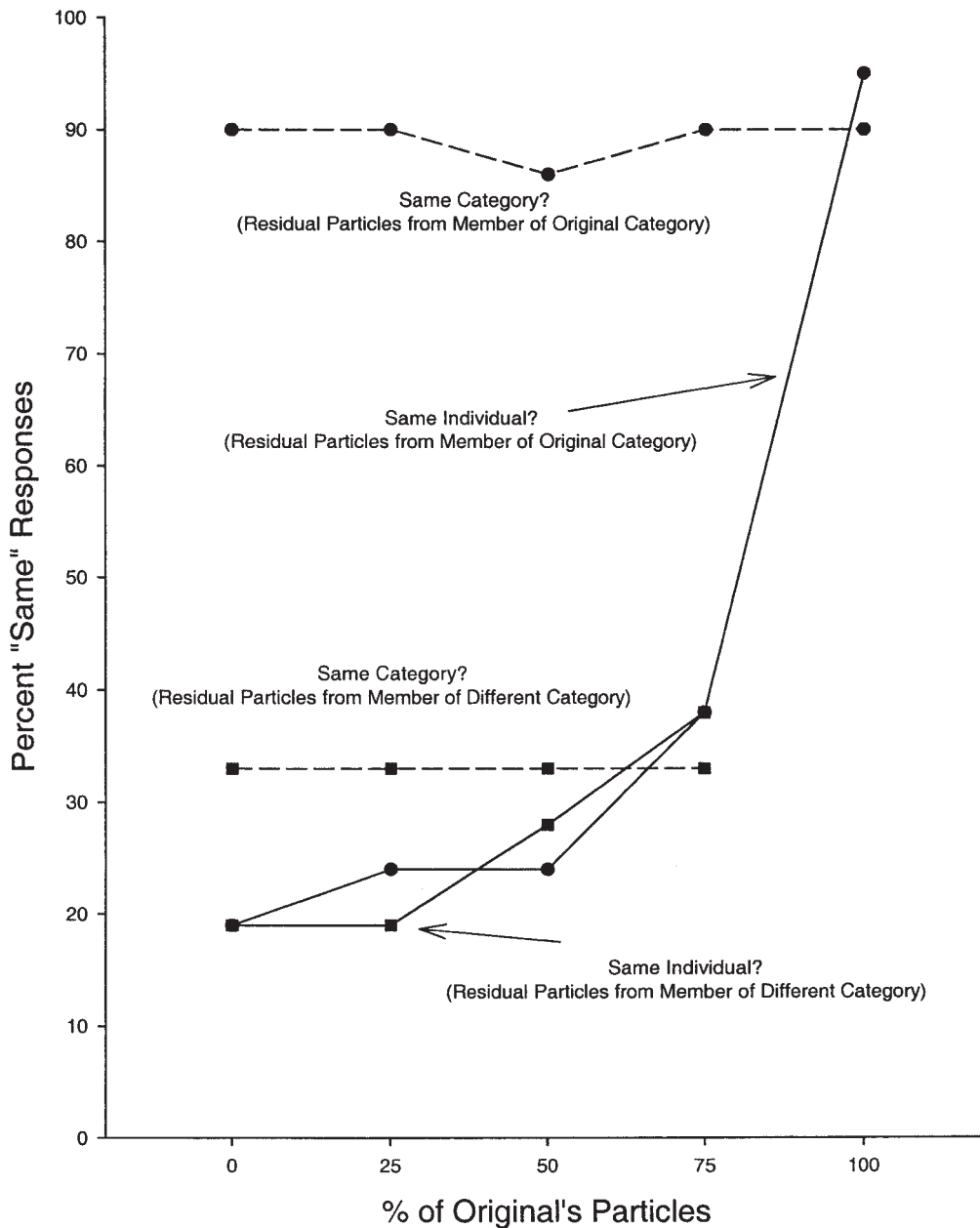


Figure 2. Percentage of responses indicating that the result of the transformation was still the same individual (solid lines) and was still a member of the same category (dashed lines). Lines with circles denote stories in which residual particles were from a member of the same category. Lines with squares indicate stories in which residual particles were from a member of a different category. For the two rightmost points, all particles came from the original object, and there are no residual particles.

particles from the original and the source of the residual particles interact with the individual versus category decision:  $Q_w(1) = 7.15, p < .01$ , for the first interaction, and  $Q_w(1) = 17.68, p < .0001$ , for the second.

Figure 2 shows that factors affecting category membership do not necessarily affect decisions about individual persistence. Although the source of the residual particles had a strong influence on category judgments, it had almost none on judgments of iden-

tity. This finding echoes the results in Figure 1 and presents another puzzle for the view that identity conditions come from knowledge of sortal membership. If *lion* is the relevant sortal, Principle 3 predicts that factors that cast doubt on whether the copy is a lion should also cast doubt on whether the copy is Fred, contrary to these results.

It is possible that category and identity judgments would have been more parallel had we used categories less similar than lions

and tigers. If lion particles had been combined with particles from robins, for example, we might have found participants less willing to accept the resulting creature as Fred. In accord with this hypothesis, both Liittschwager (1995) and Blok et al. (2005, Experiment 2) found that varying the transformational distance between the before and after states of the object decreased participants' willingness to consider it the same individual. Our claim, however, is not that change in basic-level category is always irrelevant to judgments of identity. Although we believe that causal, not sortal, factors are central to identity, category information can sometimes be informative because category changes can signal disruption of causal continuity. This is especially likely when the category shifts are across large taxonomic distances. Sortal theories, however, are committed to the idea that the meaning of sortals (their principle of identity) provides conditions for individual persistence. In particular, Principle 3 implies that if two objects are in contrasting sortals, they cannot be identical. Unless *lion* and *tiger* are not the relevant sortals, it is difficult for the sortal account to handle these results.<sup>11</sup> We return to some remaining questions about sortal theories later.

*Two-copy condition.* Data from the two-copy condition replicated those from the one-copy condition in that judgments of individual identity depended on the percentage of particles from the original individual but not on the source of the remaining particles. Figure 3 plots these results. Each small graph corresponds to a combination in which one copy contained a given percentage of particles from the original and the other copy contained another (possibly equal) percentage. The points in each graph are the percentages of trials on which participants judged: (a) that only the *dominant* copy (the one with more particles from the original) was identical to the original, (b) that only the *non-dominant* copy (the one with few particles from the original) was identical to the original, (c) that both copies were identical to the original, and (d) that neither copy was identical. When both copies have the same number of particles from the original neither copy is dominant, and we have combined the responses in which participants chose only one of these copies (these appear as *copy* responses in the graphs on the diagonal).

Several trends are highlighted. First, the percentage of dominant copy or copy responses (relative to both or neither responses) increases from top to bottom, along the columns of graphs. The increase is steep between 75% and 100% of old particles but is perceivable at lower levels as well. This indicates that as the percentage of original particles in the two copies becomes more dissimilar, participants shift toward thinking that only the dominant copy is identical to the original item. Second, a glance along the diagonal from the upper left to the lower right shows that the percentage of "both" responses increases (relative to "neither" responses). The two copies have the same proportion of original particles here, and as this proportion rises, participants increasingly believe that both copies are identical to the original. As noted earlier, it is possible to view "both" responses as reflecting intransitive judgments, and we discuss the implications of this finding later. Third, there appears to be no difference resulting from whether the residual particles came from a member of the same category as the original or from a different category. This finding replicates the results from the one-copy condition, as noted earlier. In applying the causal continuer model, we focus on these individual decisions. However, we note that decisions about category

membership in the two-copy condition (not shown in Figure 3) also replicate the one-copy condition in producing an effect of the residual particles' source but no effect of the percentage of particles from the original. This echoes the dissociation that we observed in Figure 2.

### Discussion

The causal continuer approach is consistent with these trends. According to this theory, participants' responses on a particular trial should depend on two decisions. First, they need to determine whether one of the copies is causally closer than the other. Second, they need to know whether either copy is close enough to the original to qualify as identical to it. If the answer to both questions is "yes," participants should respond that only the closer copy is identical. If the answer to the first question is "no" but the answer

---

<sup>11</sup> It is possible that participants used a concept like *mammal* or *animal* that could cover the two halves of Fred's existence and explain why they believe he is still Fred, according to a sortal account. One way of framing this hypothesis would be to say that *lion*, *mammal*, *animal*, and even *proto-object* (see footnote 3) function as sortals simultaneously, each contributing identity conditions to a particular individual like Fred who falls under them. Although Fred may no longer be a lion, he remains in enough higher level sortals to ensure his identity. This version of the theory, however, conflicts with Principle 3 that continued membership in a sortal is necessary for identity. The change from being a member to being a nonmember of any sortal category should be enough to show that the resulting object is no longer identical to the original. It is a condition of this sort that Xu and Carey (1996) implicitly appeal to in explaining their "Is-it-one-or-two?" task findings, as we noted earlier; thus, sortalists cannot abandon Principle 3 lightly. In addition, simultaneous sortals for the same object run into technical difficulties. Wiggins (2001) provides a proof that they contradict Principle 2: Leibniz's law. Another version of this hypothesis is that *lion* is the sortal for Fred under normal conditions, but participants switch to using *mammal* or *animal* as an umbrella sortal for him under the conditions that our story describes (cf. Xu, 1997). This move, however, robs the sortal theory of much of its explanatory force, because the only reason for switching sortals would seem to be the causal connection between Fred's before and after states. In general, weakening the connection between sortals and basic-level categories reduces the appeal of the sortal view by introducing vagueness about what sorts of categories can function as sortals. A sortal theory that says only that there is some category or other that provides identity information is too weak to make any serious claims. A related possibility is that there are two distinct objects that take part in the scenario, each existing simultaneously in exactly the same spatial location during the first stage of Fred's career: a lion and Fred. Each of these coinciding objects is governed by a different sortal. So if Fred continues to exist after the transformation but is no longer a lion, this simply shows that one of the original objects (i.e., Fred) survives, whereas the other (the lion) does not. Coinciding but nonidentical objects have their defenders (e.g., Wiggins, 2001), but in the present context the hypothesis shares some of the weaknesses of the umbrella theory: The only motive for positing a second sortal seems to be the details of the causal story. Furthermore, on this account, the second sortal cannot be *Fred* (since *Fred* is a proper name, not a sortal) nor *mammal* or *animal* (on pain of having to claim that ordinary lions have both a lion and a numerically distinct mammal/animal living inside them), nor *collection of particles* (because the two parts of Fred's existence involve different sets of particles in our story). What is left that avoids these problems and is not obviously ad hoc? In a later section, we return to the possibility of multiple coinciding objects but from a nonsortal perspective.

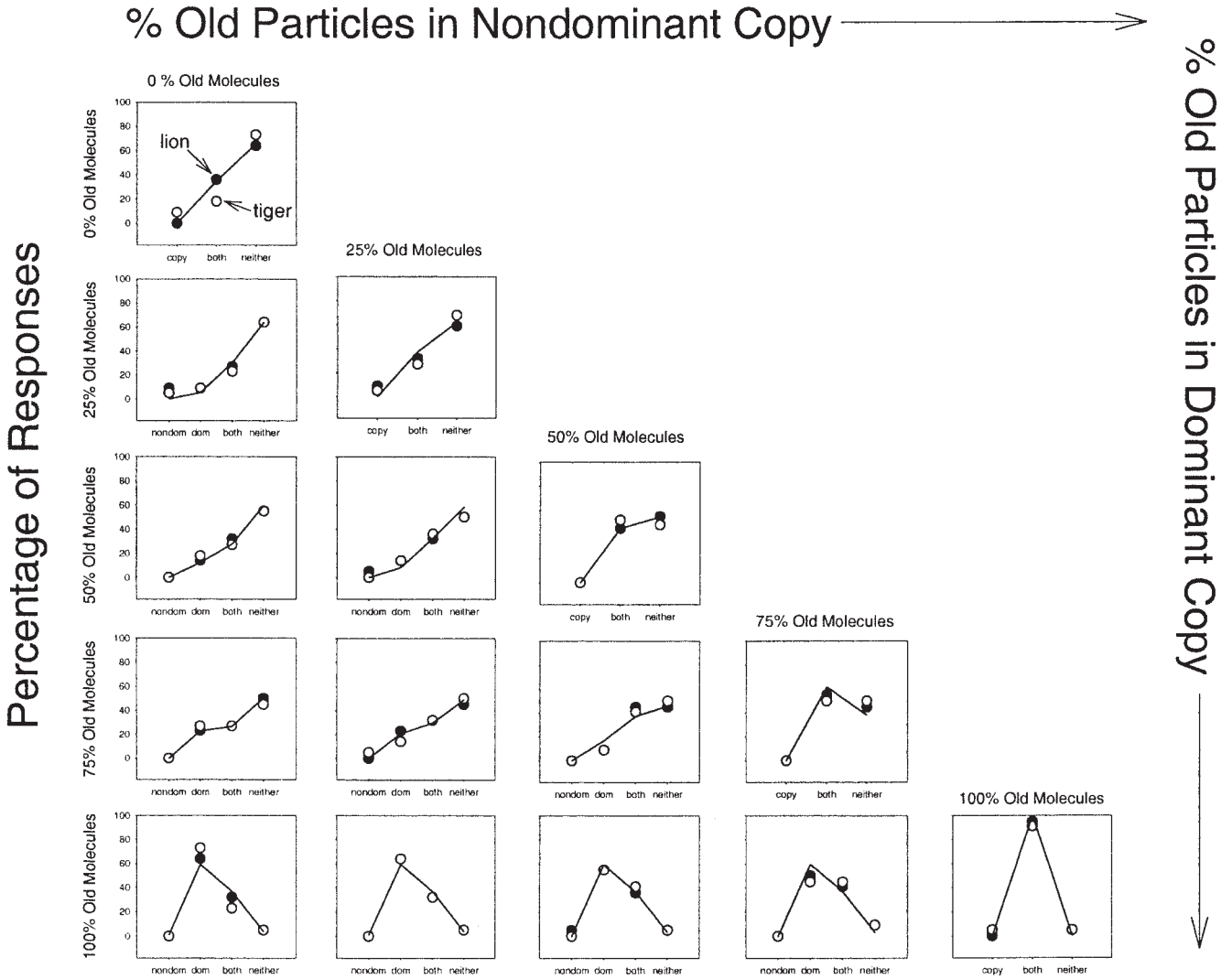


Figure 3. Percentage of responses that the dominant copy, nondominant copy, both copies, or neither copy was the same individual as the original. The graphs represents combinations in which each copy has either 0, 25, 50, 75, or 100% of its particles from the original object. Filled circles denote stories in which the residual particles were from a member of the same species. Open circles denotes stories in which the residual particles were from a member of a different species. Lines are predictions from the causal continuer theory.

to the second is “yes,” they should respond that both are identical. In all other cases (i.e., the answer to the second question is “no”), they should report that neither is identical.

We assume that causal closeness in this experiment depends on the percentage of the copy’s particles that derives from the original. In the stories, the copying machine is the causal mechanism that produces closeness by copying particles and transmitting them. We might, therefore, represent the probability that the dominant copy,  $d$ , is closer than the nondominant copy,  $n$ , in terms of the ratio in Equation 1, when the proportion of original particles in  $n$  is less than 1:

$$\Pr(d \text{ closer}) = \frac{k * (\text{Prop}(d) - \text{Prop}(n))}{1 - \text{Prop}(n)}, \quad (1)$$

where  $\text{Prop}(d)$  is the proportion of original particles in the dominant copy and  $\text{Prop}(n)$  is the proportion in the nondominant copy.

When the proportion of original particles in  $n$  is 1, we can define  $\Pr(d \text{ closer}) = 0$ . In Equation 1,  $k$  is a free parameter representing the maximum probability that  $\Pr(d \text{ closer})$  can attain. Even if copy  $d$  has all its particles from the original (i.e., the proportion of original particles in  $d$  is 1) and  $n$  has none (the proportion of original particles in  $n$  is 0), some participants might still feel that there is not enough difference between them for  $d$  to be causally closer than  $n$ . Accordingly, Equation 1 specifies that in this case  $\Pr(d \text{ closer}) = k$ . We will also assume that if  $d$  is not closer (with probability  $1 - \Pr(d \text{ closer})$ ), then we have a tie. This means that  $n$  can never be closer than  $d$  and predicts that participants will never say that only  $n$  is the original object.

The second question is whether either copy is close enough to be potentially identical to the original item. Because the same participants made identity judgments for each copy separately in the

one-copy condition, we used these earlier decisions to estimate empirically the likelihood of a “yes” answer to this question. During one of the two-copy trials, for example, participants learned that one copy contains 75% of its particles from the original and the second copy contains 25%. In the one-copy condition, participants had judged that a copy with 75% original particles was identical to the original on .38 of trials and that a copy with 25% original particles was identical on .21 of the trials. We, therefore, estimated the likelihood that one or the other is causally close enough to be identical as  $1 - (1 - .38)*(1 - .21) = .51$ . The general relationship is that in Equation 2:

$\Pr(d \text{ or } n \text{ close enough}) =$

$$1 - (1 - \Pr(d \text{ close enough})) * (1 - \Pr(n \text{ close enough})). \quad (2)$$

Combining Equations 1 and 2 gives us the predictions for the two-copy condition in Figure 3. For example,  $\Pr(d \text{ closer}) * \Pr(d \text{ or } n \text{ close enough})$  is the predicted probability that participants should identify only the dominant copy as identical to the original. Similarly,  $(1 - \Pr(d \text{ closer})) * (\Pr(d \text{ or } n \text{ close enough}))$  is the probability of a “both” response. To evaluate the model, we fit it to the data in Figure 3 using nonlinear least squares approximation. Because there is no apparent difference between cases in which the residual particles were from the same or different species, we collapsed the data from these two conditions before fitting the model. As noted earlier, the model predicts that participants should never respond that only the nondominant copy is identical to the original. Figure 3 shows that this is approximately true, but we omitted these points in fitting the model. The model was, therefore, fit to 45 data points: the “dominant only,” “both,” and “neither” responses in the 15 graphs in Figure 3. The resulting predictions appear as the lines in Figure 3, and the overall fit of the model is quite good. The root mean square deviation (RMSD) for the 45 critical observations is only 5.1 percentage points ( $R^2 = .957$ ). The value of the single free parameter,  $k$ , from Equation 1 is 0.62.

Another way to evaluate the model is to compare it with a simpler variant. Suppose, for example, that participants make their decisions on the basis of their separate judgments of whether the dominant copy is identical and whether the nondominant copy is identical. This procedure differs from the causal continuer idea in that there is no explicit comparison for closeness of the sort embodied in Equation 1. If we represent the probability that the dominant copy is close enough to be identical as  $\Pr(d \text{ close enough})$  and the probability that the nondominant copy is close enough as  $\Pr(n \text{ close enough})$ , as we did in Equation 2, then the probability that both are identical should be  $\Pr(d \text{ close enough}) * \Pr(n \text{ close enough})$ , assuming independence between the decisions. Similarly, the probability that only the dominant copy is identical is  $\Pr(d \text{ close enough}) * (1 - \Pr(n \text{ close enough}))$ , and so on. Estimating the component probabilities from the one-choice data, as we did earlier, allows us to fit this simpler model directly with no free parameters. This model does considerably less well than the one we have just described (RMSD = 16.1 percentage points and  $R^2 = .618$ ). The discrepancy is especially marked for “both” responses when the proportion of original particles is the same in the two copies, because the simpler model greatly underpredicts these proportions. In this model, a “both” response depends on both copies being independently close enough to be identical, as we just

noted. In the full model, however, there is no relevant difference between the two copies when their proportion of original particles is equal (the value of  $\Pr(d \text{ closer}) = 0$  in Equation 1); so a “both” response depends on whether either copy could be considered close enough, as given by Equation 2. This is typically a much larger value, in accord with the data. A likelihood ratio test (Bates & Watts, 1988) shows that the causal continuer model significantly improves on the simpler model, taking into account the former model’s extra parameter,  $F(1, 44) = 403.65, p < .001$ .

## Experiment 2: Causality Versus Similarity

It is reasonable to think that similarity between an object and its successor can sometimes provide evidence for identity over time. Similarity between Aunt Florence’s appearance in 1970 and in 2006 may be enough to lead us to believe that these two manifestations belong to the same person. The causal continuer model claims, however, that causal factors can override similarity if the two factors conflict. We judge someone who is merely similar to Aunt Florence, but not causally connected to her, as nonidentical, as in historical cases of imposture (e.g., Davis, 1983). To see why, imagine an iceberg whose size is  $3 \times 3 \times 3$  m at a particular time,  $t_0$ . Most people probably assume that over time icebergs tend to shrink as a result of temperature and splitting (caused by stress from storms and other factors).<sup>12</sup> Thus, at a later time,  $t_1$ , the original iceberg’s continuer would presumably have smaller dimensions rather than larger ones. The similarity of icebergs, however, might be more symmetric. For example, the  $3 \times 3 \times 3$ -m original might be about equally similar to a  $4 \times 4 \times 4$ -m iceberg and a  $2 \times 2 \times 2$ -m iceberg at  $t_1$ , but only the latter is likely to be identical to the original.

To see whether causal beliefs do indeed dominate similarity and to provide a further test of the causal continuer theory, we asked participants in a second study to make three kinds of judgments about icebergs.<sup>13</sup> In the experiment, participants read a scenario in which scientists were studying an iceberg named Sample 94, whose dimensions were  $3 \times 3 \times 3$  m. During the first two parts of the experiment, we gave participants a list of icebergs of varying dimensions (e.g.,  $4 \times 3 \times 1$  m or  $2 \times 1 \times 1$  m) that the instructions described as being found sometime later in the same vicinity. Participants rated how similar each item was to the original Sample 94 and also how likely it was that the item was Sample 94. Participants made their ratings by circling a number on a scale ranging from 0 to 9. For similarity ratings, the endpoints had the labels *extremely dissimilar* (0) and *extremely similar* (9); for the identity ratings, the endpoints had the labels *extremely unlikely* (0) and *extremely likely* (9).

In the third part of the experiment, participants read that scientists had spotted pairs of icebergs in the same vicinity as Sample 94. Participants saw a list of pairs, with one item in each pair labeled *A* and the other labeled *B* (e.g., Iceberg *A* is  $2 \times 1 \times 1$  m and Iceberg *B* is  $3 \times 3 \times 2$  m). For each pair, participants made a forced choice among the alternatives: “Only Iceberg *A* is Sample 94,” “Only Iceberg *B* is Sample 94,” “Both Icebergs *A* and *B* are Sample 94,” and “Neither Iceberg *A* nor Iceberg *B* is Sample 94.”

<sup>12</sup> This assumption is also factually correct. Although it might seem that icebergs would have to grow before they can shrink, in fact icebergs are created when they break off from ice shelves in Arctic or Antarctic regions.

<sup>13</sup> We thank Douglas Medin for suggesting this idea.

The goal of the first two parts of the study was to distinguish identity and similarity judgments. If causal mechanisms dominate judgments of identity, we should find that participants give lower identity ratings than similarity ratings to icebergs whose dimensions are greater than the original sample. Similarity and identity judgments may converge for icebergs whose dimensions are smaller than the original. The goal of the third part of the study was to determine whether the causal continuer model could predict participants' decisions about which iceberg in each pair (or both or neither) was identical to the initial one. We assume these decisions will again reflect the model's two-part structure: the participants' notion of whether either alternative is causally close enough to be the original and also whether one alternative is causally closer than the other.

**Method**

*Stimulus items.* In their similarity and identity ratings, participants compared the  $3 \times 3 \times 3$ -m iceberg to each of a set of items that we formed by combining the dimensions 4 m, 3 m, 2 m, and 1 m in all distinct ways. Thus, one item was  $4 \times 4 \times 4$  m, another was  $4 \times 4 \times 3$  m, and so on. The instructions told participants that the dimensions were always given

with the larger sides first without regard for the iceberg's orientation. For example, participants rated a  $4 \times 3 \times 1$ -m iceberg but not a  $3 \times 1 \times 4$ -m iceberg, because these would be considered the same item. Because of this aliasing, there were a total of 20 items in the stimulus set (shown on the x-axis of Figure 4). The similarity booklet listed these in a new random order for each participant. A rating scale followed each item that contained the numbers 0 to 9 on a horizontal line, and the participant circled a number to indicate the rating. The first page of the booklet contained the instructions; the following two pages contained the items to be rated. We constructed the identity booklets in the same way, except for the change in instructions and the labeling of the response scale. The two booklets for a given participant listed the items in a different random order.

We selected a subset of the items from the similarity and identity ratings to create the iceberg pairs for the final part of the experiment. We discarded pairs in which the two icebergs within the pair would not simultaneously fit inside Sample 94's original  $3 \times 3 \times 3$ -m shape. For example, one item paired a  $2 \times 2 \times 2$ -m iceberg with a  $1 \times 1 \times 1$ -m iceberg, but there was no pair with a  $2 \times 2 \times 2$ -m and a  $3 \times 3 \times 2$ -m iceberg. This left 39 pairs to be tested. Within each pair, we randomly designated one item as Iceberg A and the other as Iceberg B. Beneath each pair in the booklet were the four response choices (i.e., only A, only B, both, and neither in that order), and participants made their decision by circling one of these options.

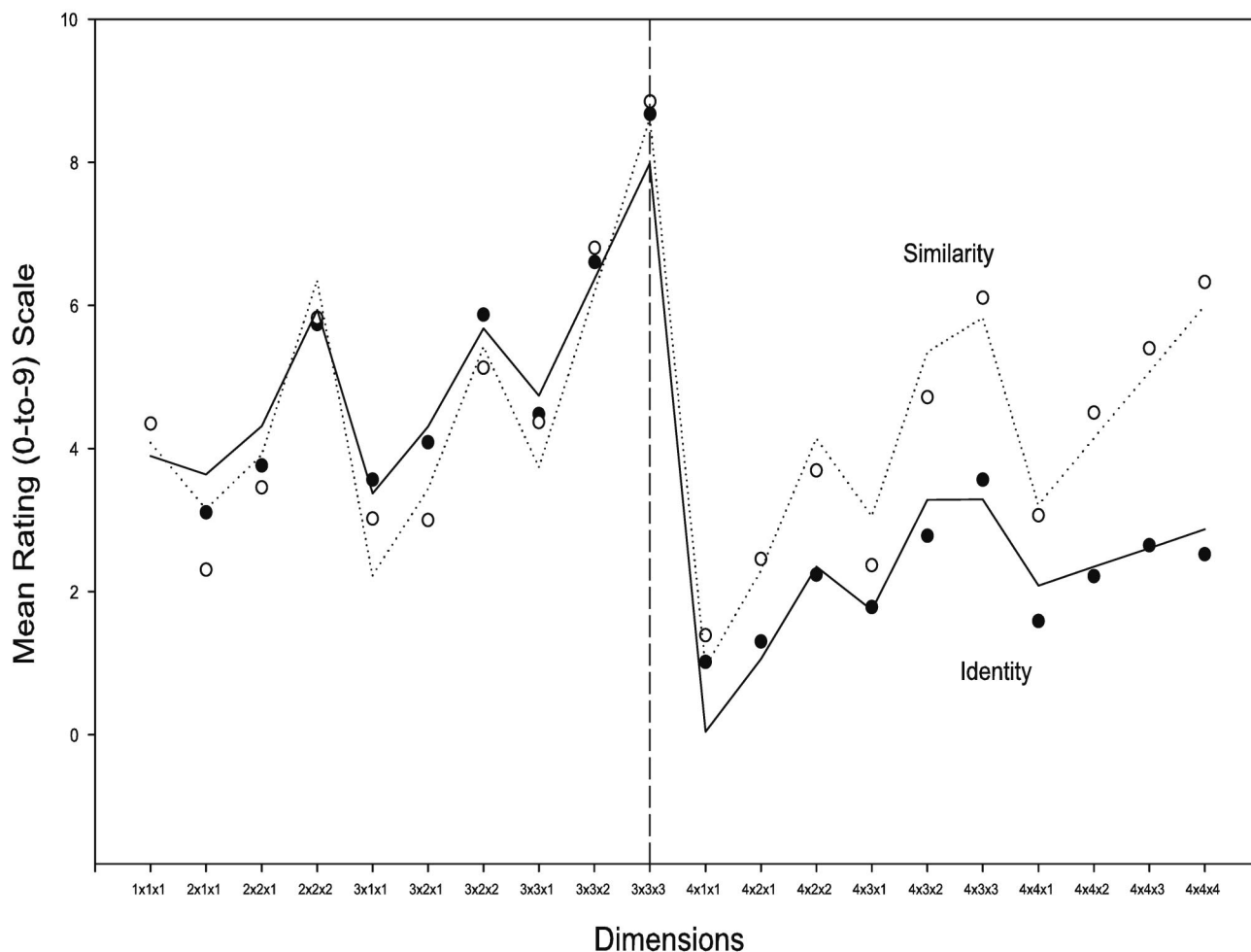


Figure 4. Mean ratings (0-to-9 scale) of similarity (open circles) and identity (filled circles) between icebergs of varying sizes (x-axis) and a  $3 \times 3 \times 3$  standard. The dotted line shows predictions for similarity ratings from a regression model; the solid line shows predictions for identity ratings (see text for a description of the models).



*Participants.* The participants were 46 Northwestern University undergraduates who were enrolled in introductory psychology and participated for course credit. None had been in Experiment 1. Half the participants rated similarity first; half rated identity first.

## Results

*Identity and similarity ratings.* We predicted that when comparing the standard iceberg ( $3 \times 3 \times 3$  m) with one with a larger dimension (e.g.,  $4 \times 3 \times 3$  m), participants would tend to see the second as potentially similar, but not identical, to the first. Because icebergs tend to shrink over time, a comparison iceberg with a larger dimension can be similar but not identical to the standard. The mean ratings appear in Figure 4, and they confirm this prediction. When the comparison iceberg has a larger dimension (right side), mean similarity ratings are always higher than identity ratings, but when the comparison iceberg's dimensions are smaller or equal to those of the standard, the ratings tend to be more nearly equal (left side).

An analysis of variance (ANOVA) on these factors shows that there is a significant interaction between type of rating (similarity vs. identity) and whether the iceberg has a dimension greater than that of the standard,  $F(1, 45) = 40.91$ ,  $MSE = 12.95$ ,  $p < .0001$ . It is more revealing, however, to examine variables that may have contributed to participants' reasoning about these judgments. Figure 4 shows peaks in the ratings when the icebergs were cubical (e.g.,  $2 \times 2 \times 2$  or  $4 \times 4 \times 4$ ) or nearly so (e.g.,  $3 \times 2 \times 2$  or  $4 \times 3 \times 3$ ), suggesting that participants were taking into account the iceberg's shape. Because the standard iceberg was itself cubical, participants may have given the comparison iceberg higher ratings if it too had approximately the same shape. In addition, participants seem to have taken overall size into account, giving higher ratings when the size of the comparison iceberg was near that of the standard. Initial analyses hinted that participants may have compared the icebergs in terms of the sum of their dimensions rather than the product, possibly for computational ease. Icebergs whose dimensions summed to near 9, the sum of the dimensions of the  $3 \times 3 \times 3$  standard, tended to get higher identity and similarity ratings than the others. Compare, for example, the ratings of the  $4 \times 3 \times 2$  item (dimension sum = 9) with the  $4 \times 2 \times 1$  item (dimension sum = 7) in Figure 4.

To see how well these factors predicted the mean ratings, we applied two regression equations to the Figure 4 data, one for the similarity and the other for the identity judgments. Both equations contained three terms: The first captured departure from cubical shape in terms of the standard deviation of the iceberg's three dimensions. The second measured the overall difference in size between the standard and comparison iceberg using the sum of the dimensions, as we have just discussed. That is, if  $d_1$ ,  $d_2$ , and  $d_3$  are the dimensions of the comparison iceberg, then the value of this term was  $|d_1 + d_2 + d_3 - 9|$ . The final term was a binary indicator of whether any of the iceberg's dimensions was greater than that of the standard (1 if one or more of the dimensions was 4 m, and 0 otherwise). We expected this last term to discriminate the identity ratings from the similarity ratings. Predictions from these two regression equations appear in Figure 4: The solid line corresponds to the identity predictions and accounts for 95.4% of the variance among the means. All three terms produced statistically significant coefficients: for the shape term,  $b = -1.63$ ,  $SE = 0.209$ ,  $t(16) = 7.80$ ; for the size term,  $b = -0.68$ ,  $SE = 0.072$ ,  $t(16) = 9.47$ ; and for the indicator term,  $b = -3.07$ ,  $SE = 0.253$ ,  $t(16) = 12.12$ ;  $p < .0001$  in all cases. Figure 4 shows the predictions for the similarity

ratings as the dotted line, and these predictions account for 92.6% of the data. The shape and size terms were significant in this analysis: for shape,  $b = -2.92$ ,  $SE = 0.251$ ,  $t(16) = 11.62$ ; for size,  $b = -0.76$ ,  $SE = 0.086$ ,  $t(16) = 8.76$ ;  $p < .0001$  in both cases. However, as we expected, there was no effect on the similarity ratings of whether the comparison iceberg contained a dimension larger than those of the standard:  $b = -0.35$ ,  $SE = 0.304$ ,  $t(16) = 1.16$ ,  $p = .26$ . When one of the comparison iceberg's dimensions exceeds that of the standard iceberg, participants discount the possibility that it could be the standard but not the possibility that it could be similar to it.

A possible objection to these results is that task demands may have pressured participants to respond differently when rating similarity than when rating identity. Because a participant made both types of responses, he or she may have felt that the instructions pragmatically implied that the two sets of ratings should differ in some important way. In this study, though, we blocked the two sets of ratings (unlike Experiment 1 in which identity and category judgments were interleaved). Thus, we can analyze the first block of ratings a participant made. During this first block, participants did not know about the subsequent rating task, and their responses should, therefore, be free of the pragmatic influence. An ANOVA of these first-block data, however, shows the same interaction between type of rating and size of the comparison iceberg that we reported earlier,  $F(1, 44) = 17.02$ ,  $MSE = 19.33$ ,  $p < .001$ .

## Discussion

*Extensions of the causal continuer model for the data of Experiment 2.* We can apply the causal continuer theory to predict the results of the final part of Experiment 2, in which participants decided for each of a series of pairs whether one, both, or neither item of the pair was the original iceberg, Sample 94. The response distributions for each stimulus pair appear in Figure 5 in a format similar to that of Figure 3. For purposes of displaying the data, we have changed the A and B labels of the icebergs so that all graphs would fall on or below the diagonal; the original assignment of labels was random. Empty cells in Figure 5 correspond to pairs that would not have fit within the original dimensions. For example, it is not possible for a  $2 \times 2 \times 2$  solid and a  $3 \times 2 \times 2$  solid to occupy simultaneously a  $3 \times 3 \times 3$  space. As mentioned earlier, we did not test these pairs.<sup>14</sup>

<sup>14</sup> We might expect the predictions to be more complex than in Experiment 1, because of possible dependencies between the paired items. For example, the likelihood that a  $3 \times 3 \times 2$ -m iceberg is Sample 94 might be greater if the other iceberg in the pair is  $2 \times 2 \times 2$  than if it is  $1 \times 1 \times 1$ . (A  $3 \times 3 \times 2$  and a  $1 \times 1 \times 1$ -m iceberg will both fit inside a  $3 \times 3 \times 3$  outline, but a  $3 \times 3 \times 2$  and a  $2 \times 2 \times 2$ -m iceberg will not.) Experiment 1 minimized these dependencies by creating two independent copies of each original, but Experiment 2 included no such duplicating process. Dependencies of this sort between potential continuers are not contrary to the causal continuer theory, because decisions about relative closeness could take into account the full set of continuers. However, these dependencies would call for a more complex set of mathematical assumptions than those we have dealt with so far. As we noted earlier, we attempted to reduce the dependencies by including a pair only if the members of the pair could both fit within the original  $3 \times 3 \times 3$  shape. Although dependencies are still possible, it is worthwhile determining whether a simple version of the model is able to capture the main features of the data.

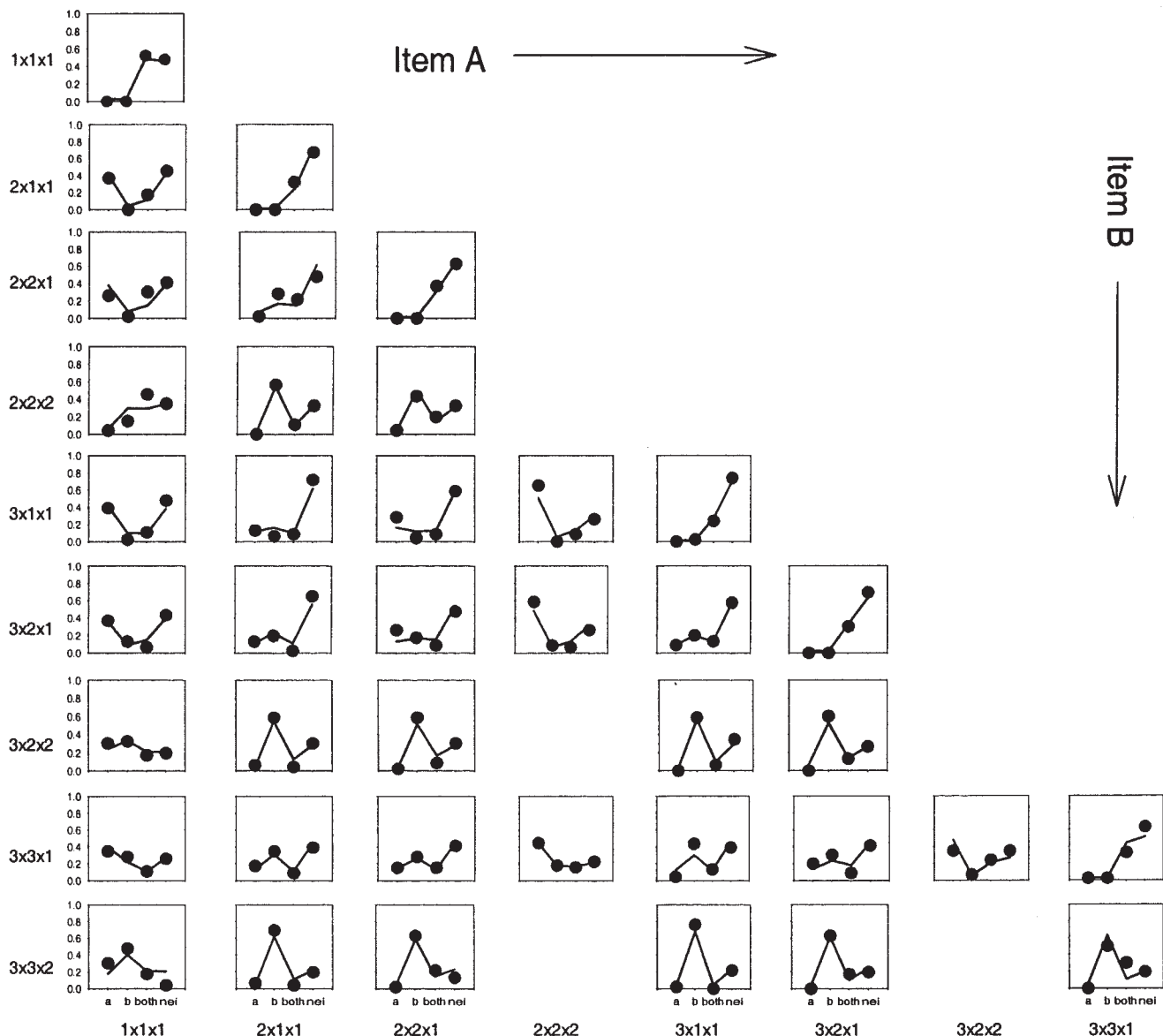


Figure 5. Percentage of responses that Iceberg A, Iceberg B, both icebergs, or neither iceberg is identical to the original. Each of the small graphs represents a pair of icebergs whose constituents are shown on the corresponding row and column. Lines are predictions from the causal continuer theory.

We note first that the data (circles) are generally consistent with the results from the first part of the experiment in that icebergs with dimensions that received high identity ratings when considered alone tend also to be selected as the continuer when paired with a less highly rated item. For instance,  $2 \times 2 \times 2$ ,  $3 \times 2 \times 2$  and  $3 \times 3 \times 2$  items all tended to receive high identity ratings individually (see Figure 4) and also to dominate the pairs that include them (see Figure 5). Likewise, “neither” responses tended to be more frequent when both members of the pair had relatively low identity ratings. When a  $3 \times 1 \times 1$  iceberg is paired with a  $2 \times 1 \times 1$  iceberg, for example, participants often responded that neither can be the original. There are some exceptions to this pattern of consistent responding, however, which we consider later.

In fitting the causal continuer theory to these data, we do not have a simple, objective measure of closeness in the way we did in Experiment 1: As we have seen, identity ratings in this study depended on a mix of overall shape and size. For this reason, we cannot apply Equation 5 directly to determine whether one iceberg is closer than the other. However, it is possible to obtain a subjective measure of closeness by using the identity ratings themselves from the first part of the experiment. In particular, suppose that on a given trial participant  $i$  perceives Iceberg A to be close to Sample 94 to degree  $x_{iA}$  and Iceberg B to be close to Sample 94 to degree  $x_{iB}$ . We assume that  $x_{iA}$  and  $x_{iB}$  are drawn from a normal distribution whose mean is equal to participant  $i$ 's identity ratings for A and B,  $A_i$  and  $B_i$ , respectively, and whose standard deviation

is  $s$ . Suppose, too, that participant  $i$  deems Iceberg A closer than Iceberg B if the difference between these values,  $x_{iA} - x_{iB}$ , is greater than some criterion value,  $d$ . The probability of this occurrence is given by Equation 3, where  $N$  is the normal density function, with mean equal to the difference between the ratings for A and B:

$$\Pr(A \text{ closer}) = \int_d^\infty N(A_i - B_i, \sqrt{2}s) dy. \quad (3)$$

A similar expression with  $A_i$  and  $B_i$  reversed gives the probability that the participant will judge B closer than A. The probability of a tie, neither A nor B is closer, is then  $1 - \Pr(A \text{ closer}) - \Pr(B \text{ closer})$ . This information is parallel to what Equation 1 provides for Experiment 1.

To determine whether either A or B is close enough to Sample 94, the participant compares  $x_{iA}$  and  $x_{iB}$  with a second threshold value,  $v$ . Equation 4 then expresses the probability that either item exceeds this value, under the same assumptions mentioned earlier:

$$\Pr(A \text{ or } B \text{ close enough}) = 1 - \left( 1 - \int_v^\infty N(A_i, s) dy \right) * \left( 1 - \int_v^\infty N(B_i, s) dy \right). \quad (4)$$

This yields information comparable to that in Equation 2. Combining Equations 3 and 4, as before, produces the predicted probability that only Iceberg A, only Iceberg B, both, or neither is Sample 94. In fitting the model, we used only the proportion of “only A,” “both,” and “neither” responses. The proportions of “only B” responses were obtained by subtraction. Figure 5 displays these “only B” responses for reference, along with the others, although they are not included in the goodness-of-fit statistics. We treated the standard deviation,  $s$ , and the threshold values,  $d$  and  $v$ , as the same for each participant and estimated them using the technique utilized in Experiment 1. For this model,  $\text{RMSD} = 6.5$  percentage points and  $R^2 = .888$ . This fit is somewhat less close than that of Experiment 1, perhaps in part because of the larger number of points that the model must predict. The best fitting value for  $d$ , the difference needed to judge one iceberg closer than the other, was 0.91 points on the 0–9 scale, and the value of  $v$ , the threshold value for an iceberg to be close enough to be Sample 94, was 5.31 on the same scale. The standard deviation,  $s$ , was 0.39 points.

It is again instructive to compare the model with a simpler variant that posits only a single type of comparison. Participants could easily accomplish their task of deciding which of the icebergs is Sample 94 by determining separately whether Icebergs A and B are above the threshold value  $v$ . They could then respond that only A is Sample 94 if A but not B is above threshold, that both are Sample 94 if both are above threshold, and so on. This leaves out the comparison between A and B that Equation 3 captures. We have fit this type of standard threshold model to the data of Figure 5, using distribution assumptions similar to those introduced in connection with Equation 4. As in Experiment 1, the simpler model yields a less convincing account of the results,

although the difference between the models is not as dramatic as in the earlier experiment. For the simpler model,  $\text{RMSD} = 9.0$  and  $R^2 = .796$ ; the obtained values of  $v$  (5.64) and  $s$  (0.90) are approximately the same as in the original model.

It is possible to compare predictions to the data from individual participants to explore reasons for the model’s successes and failures. The causal continuer model (but not its simpler variant) is consistent with the possibility that a participant could give above-threshold ratings to a particular Iceberg A ( $x_{iA} > v$ ) but then choose an even more highly rated Iceberg B as Sample 94 (rather than responding “only A” or “both”) when A and B are paired. This is in line with the model’s context sensitivity, which we noted earlier: Even when both icebergs are above threshold, “only B” is the predicted response if B is closer than A. This was a fairly common occurrence in the data. Among the 467 cases in which the ratings for both icebergs exceeded the obtained threshold value of 5.31, only the dominant iceberg was picked as Sample 94 on 28% of trials in preference to a “both” response. What are more difficult for the model to accommodate are cases in which one iceberg dominates the other in a participant’s ratings, but he or she nevertheless responds “both” or selects the dominated item when the icebergs are paired. In the 1,259 cases in which one iceberg was rated more highly than the other, this type of switch occurred on 24% of trials. Although some of these cases may reflect noise in the ratings and can be handled by the distributional assumptions that we have made in Equations 3 and 4, inspection of the data suggests that some of the violations may be due to systematic changes in the participants’ emphasis on shape versus size of the icebergs, as they moved from rating single items to making paired comparisons. This finding may echo similar kinds of reversals in the decision-making literature (see, e.g., Shafir et al., 1993). Although it would be possible to reformulate the model to take into account such shifts between properties, we have not attempted to do so here.<sup>15</sup>

*Extensions and limitations.* Although we believe that the causal continuer model has some advantages over earlier approaches, we have considered so far only a fairly narrow range of identity judgments. We have focused on situations that are difficult for other theories to explain by deliberately eliminating spatiotemporal continuity (e.g., destroying the particles of the original object in the scenarios of Experiment 1) and by dissociating identity from basic-level category membership (Experiment 1) and similarity (Experiment 2). It is worth asking, however, whether the model

<sup>15</sup> We note, too, that it is theoretically possible in our formulation of the causal continuer model for participants to judge that one of the icebergs, for example, Iceberg A, is close enough to be Sample 94, that Iceberg B is not close enough but that B is closer than A. This stems from the simplifying assumption that judgments about whether A or B is close enough are independent of judgments about whether A or B is closer. Computations based on the parameter values we derived earlier show, however, that the model predicts these inconsistencies to be rare. The predicted probability of such cases is greater than 1% of trials only when both icebergs receive a rating of 5, and in this situation the predicted probability is only 1.7%. (We cannot check these predictions against the data, because we have no independent way of determining the probability that a participant judged each iceberg close enough from his or her ratings.) It is possible to eliminate the independence assumption but at the cost of a more complex model (see also footnote 14).

deals with other types of identity judgments and with special properties of the identity relation. In the following section, we examine some additional findings to see how well the theory generalizes. We begin by considering a further experiment that explored identity decisions that may be closer to real-world situations than those discussed earlier. We then consider some potential challenges to the causal continuer approach concerning domain specificity, sortal status, and transitivity of identity.

### Experiment 3: The Identity of Rivers and Streets

Experiments 1 and 2 use situations that are similar to the ship of Theseus story in that a single object divides into two potential continuers. Temporal fission of this sort may be fairly rare, but there are analogous spatial situations that are quite common. One situation that approximates naturally occurring fission is branching along routes, such as rivers or roads. These cases have served as an important source of evidence and analogy for theories of identity (e.g., Chisholm, 1976; Lewis, 1983; Sider, 2001). For example, Lewis (1983) argues that the identity relation is not always decisive in determining how to count objects, because we can count nonidentical roads (e.g., the Chester A. Arthur Parkway and Route 137) as a single road in a stretch where they briefly merge. If the causal continuer theory can account for intuitions about routes, it may clarify the basis of such appeals. For the same reason, failure to explain these cases might be taken as evidence against the theory.

Imagine, then, that you are canoeing up the Green River searching for a campsite that you know lies along that river, and you suddenly encounter a fork. Which way should you turn to stay on the Green River? Both branches are causally connected to the main part of the river, but it is possible that the connection is closer for one branch than the other. For example, if the size of one branch is larger than the other, then the connection might be more direct for the larger branch, and it might, therefore, be a better candidate for the river's continuation. Similarly, if the angle of one branch is closer to dead ahead than the angle of the other, it might be a better candidate.

Of course, there are differences between determining the identity of objects over time (and space) and determining the continuation of routes over space. The causal factors that relate the main trunk of a river to its branches are different, and perhaps less direct, than the factors that relate an object to its temporal successors. Nevertheless, there is a connection between these cases. The flow of water between the branches and the trunk of a river can be seen as a causal link, with the width and angle of the branches indexing the relative strength of this link. There is no doubt that other factors, such as the underlying terrain of a river basin, also affect the form the river takes, but environmental factors also play a role in cases of temporal branching, as when wind or waves cause splitting of icebergs. Even in the case of man-made routes such as streets, where there is no obvious causal force that the trunk exerts on its branches, historical causes may substitute for synchronic ones. Many streets originated from vehicle or pedestrian traffic from one landmark to another, and these historical forces may leave their mark on the angles and widths of the branches. Much the same may be true even of contemporary streets, where mechanical equipment carves out the path that the street will take. Our present goal is to show that the same sorts of

comparisons that the causal continuer theory uses to explain temporal cases can also predict people's judgments about spatial fission.

What predictions can we draw from the causal continuer model for this type of decision? In Experiment 3, we will not attempt to predict all logically possible responses, as we did in Experiments 1 and 2, but will attempt to predict changes in specific responses across conditions. We focus on the case in which one branch, the other, or neither must be the continuation. According to the model, "neither" responses can come about for two reasons: First, if there is no relevant difference between the two branches (neither is closer than the other), then neither can claim to be the original. In the canoeing example, both branches may have roughly the same angle from dead ahead, roughly the same width, and so on. Second, both branches may be too far from the original to qualify as a plausible continuer (neither is close enough). This might occur, for example, if even the wider of the Green River's two branches is quite narrow and if even the branch with the angle closer to dead ahead still makes quite a large angle with the main stream. It is difficult to test these predictions using actual rivers and roads because many factors affect the naming patterns of these routes: not only angle and width, but also length, traffic, current flow, and the like. However, we may be able to check the predictions in a simplified situation in which we can control irrelevant variables.

We noted that the causal forces that govern streets may be less direct than those governing rivers, in line with the status of streets as artifacts and rivers as natural kinds. If so, we might expect a larger number of "neither" responses for streets than rivers. A "neither" response in this context means that the Green River or Green Street has come to an end and two new rivers or streets have begun at the fork. This abrupt stopping and starting may be more difficult to achieve for rivers because of the stronger causal continuity inherent in the flow of water from branch to trunk.

### Method

*Stimuli.* To look at these possibilities experimentally, we asked participants to judge which branch of a schematic river or street was the continuation of a labeled initial segment. On each trial, participants saw an inverted Y-shaped figure on a computer screen (see the superimposed diagrams in Figures 6 and 7). The branches of the Y varied independently in their angle from the vertical and in their width. The angles were 0°, 30°, 60°, and 90° from vertical; the widths were 1, 2, 3, and 4 mm. The trunk of the inverted Y was always vertically oriented, and its width was the sum of the widths of the branches. Each combination of values of angle and width for the right branch appeared with each combination of values for the left, except for the impossible case in which both branches were at 0°. The branches' lengths were always equal.

*Procedure.* In one condition, we told participants that they would be viewing possible maps of a river, that the top part of the inverted Y was called the Green River, and that they should decide for each map which branch (if either) was the Green River's continuation. A second condition was similar, except that the instructions told participants that the maps represented Green Street. A computer presented the trials in a new random order for each participant. The participants selected the left branch as the continuation (by pressing the *F* key on their keyboard), the right branch (by pressing the *J* key), or neither branch (by pressing the space bar).

*Participants.* There were 20 participants in each condition. They were from the same pool as those in the earlier studies but had not taken part in those experiments.

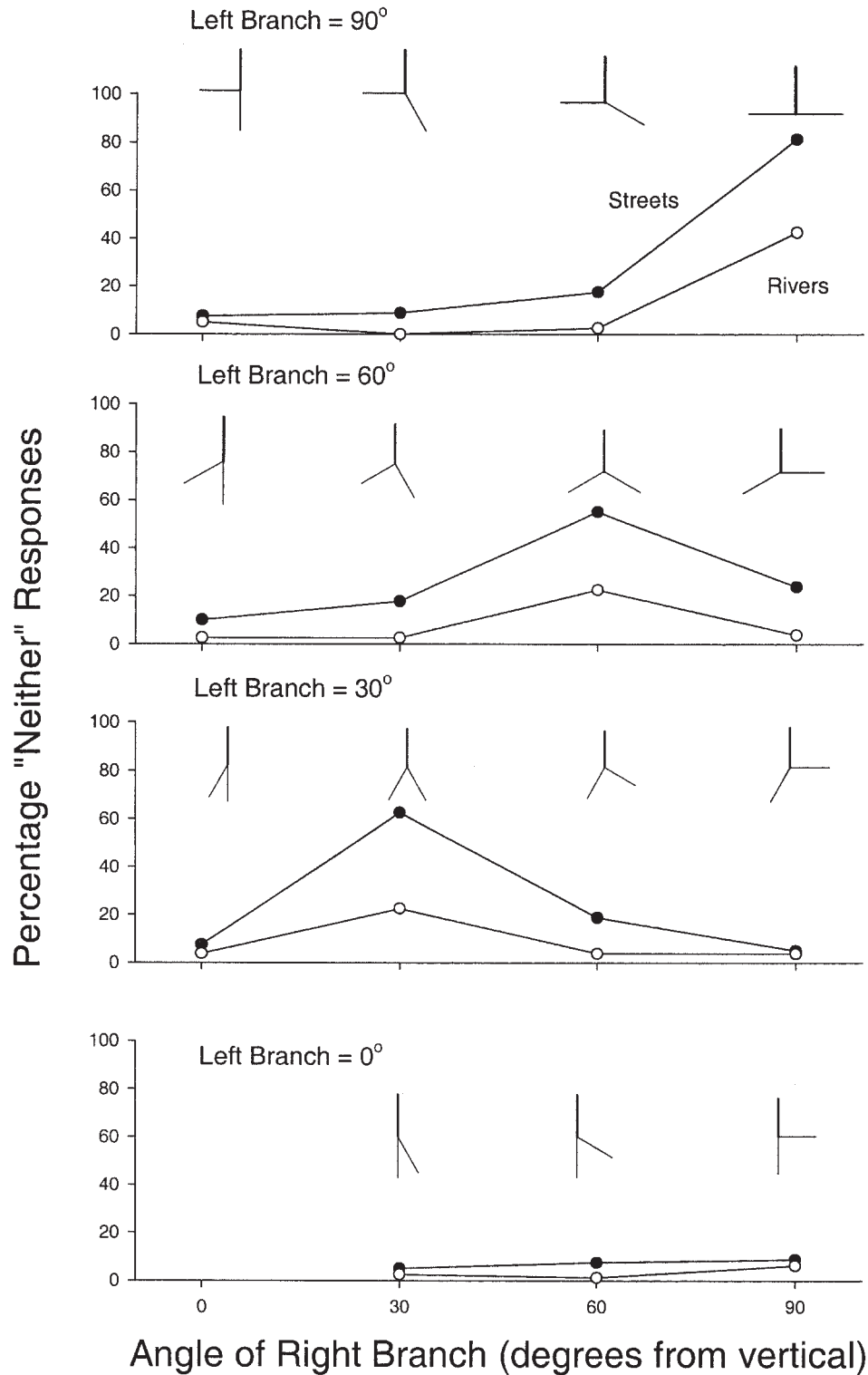


Figure 6. Percentage of trials in which participants responded that neither branch was the continuer of the initial portion of the street or river. Results appear as a function of the angle between each branch and the vertical. Filled circles represent the streets condition, and open circles the rivers condition. Superimposed diagrams indicate a sample stimulus item associated with the data point below it. (Data are taken from only those trials in which the widths of the two branches are equal.)

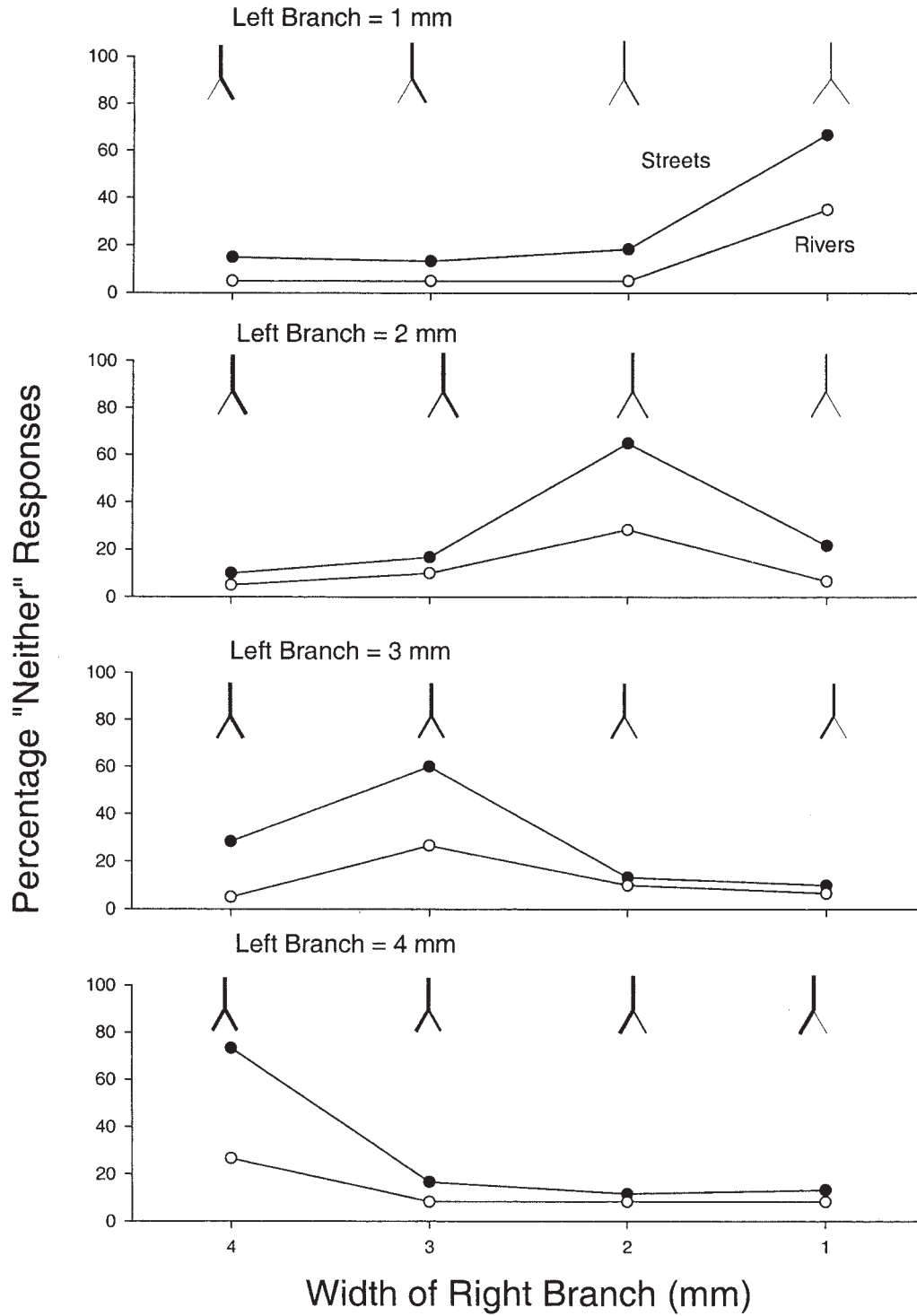


Figure 7. Percentage of trials in which participants responded that neither branch was the continuer of the initial portion of the street or river. Results appear as a function of the width of each branch. Filled circles represent the streets condition, and open circles the rivers condition. Superimposed diagrams indicate a sample stimulus item associated with the data point below it. (Data are taken from only those trials in which angles of the two branches are equal.)

## Results and Discussion

In both the streets and the rivers conditions, participants' choice of the right or left branch depended on the width and angle of the branch. For example, participants' choice of the right branch as the continuer increased from 30% to 68% in the street condition and from 28% to 67% in the rivers condition as that branch straightened from 90° to 0°. This difference resembles in some respects variations in the reliability of edges in studies of illusory contours (Kellman & Shipley, 1991). Similarly, choice of the right branch increased from 22% to 68% in the streets condition and from 23% to 74% in the rivers condition as that branch thickened from its narrowest to its widest value.

The more interesting results are the "neither" responses. We predicted, first, that these responses should increase as the angle and width of the two branches approach each other. This difference effect is easiest to spot in the data if we hold the width of the two branches constant while inspecting the angle differences and, similarly, hold angles constant while inspecting the width differences. The results for angles appear in this conditionalized form in Figure 6. The individual graphs represent the angle of the left branch with respect to vertical, and the  $x$ -axis in each graph shows the angle of the right branch. The diagram above each point in Figure 6 illustrates a sample stimulus item that contributes to that point. When the angles are equal (the diagonal points from the lower left to the upper right of the figure), "neither" responses peak. The "neither" responses decrease systematically as the difference between the angles increases. (There are no data at the 0° point in the bottom graph because the branches cannot both be vertical simultaneously; for the remaining points in this graph, participants overwhelmingly chose the straight branch as the continuer.) To evaluate this effect, we conducted a multiple logistic regression (Hosmer & Lemeshow, 1989), in which the dependent variable was whether a participant had made a "neither" response and the independent variables were the difference between the angles of the branches and the size of the smaller angle (with respect to vertical). In both the streets and rivers conditions, the results in Figure 6 show a significant effect for the angle difference:  $Q_w(1) = 68.51$  for streets and  $Q_w(1) = 22.66$  for rivers;  $p < .0001$  in both cases. We describe the effect of the smaller angle shortly.

Figure 7 contains the comparable data for the width of the branches. (We have reversed the scale on the  $x$ -axis to make the results easier to compare with those of Figure 6. Each Y-shaped diagram illustrates one of the stimulus items associated with the point below it.) The width data show a similar pattern of decreasing "neither" responses as the difference between the widths increases:  $Q_w(1) = 112.62$  for streets and  $Q_w(1) = 33.54$  for rivers;  $p < .0001$  in both cases.

The second prediction from the causal continuer theory was that participants should also respond "neither" if the closest of the two branches is not close enough to the original: a "min effect." In the case of the branch angles in Figure 6, we should expect that as the smaller of the two angles increases (gets farther from vertical), the percentage of "neither" responses should increase. To see this effect in the streets data, notice, for example, that when the left branch is at 0° and the right branch is at 30° (bottom graph of Figure 6), only 5% of responses are "neither." However, this response rate increases to 19% when the left branch is at 30° and

the right branch is at 60° (third graph from the top), and it increases further to 24% when the left branch is at 60° and the right branch is at 90° (second graph from the top). In each of these cases, the difference between the angles is constant at 30° (i.e., 0° vs. 30° from vertical, 30° vs. 60°, and 60° vs. 90°); so the increase is independent of the difference effect just discussed. Overall,  $Q_w(1) = 22.93$ ,  $p < .001$  for streets, and  $Q_w(1) = 9.95$ ,  $p < .01$ , for rivers. For the width variations, however, no such effect occurred:  $Q_w(1) = 0.39$  for streets and  $Q_w(1) = 0.50$  for rivers;  $p > .10$  in both cases. It is possible that even our smallest width value (1 mm) was large enough to be a possible continuer for our items because it was clearly visible in the stimulus displays.

In line with our final prediction, "neither" responses are uniformly lower for rivers than for streets. Participants are apparently more willing to suppose that a street can end abruptly and two new streets continue from that point (so that a "neither" response is appropriate) than that a river can end and two new rivers begin. Participants might believe, for example, that authorities are more likely to redesignate streets at their convenience than to redesignate rivers.

These results suggest that the causal continuer model may apply to identity across space as well as over time. The model confers identity based on whether an item is close enough to be a bona fide continuer and on competition among these continuers. It, therefore, predicts both a min effect (the identical item must have minimal qualifications) and a difference effect (the identical item must be better than other qualified items). In line with these predictions, we found both a difference effect and a min effect for angles and a difference effect for widths in both the streets and rivers conditions. The results of the experiment, then, provide some generality for the model.

## General Discussion

### *The Issue of Domain Specificity*

The difference between streets and rivers that appears in Figures 6 and 7 raises an important issue about the nature of the causal continuer theory. The theory assumes that (people believe that) causal forces and the objects they create exist in their own right, independent of language and thought. In particular, physical objects do not depend on the concepts or categories to which these objects belong. In this sense, the theory is domain general, because the theory applies causal closeness as a metric to physical objects of all sorts. However, how can the theory account for the evident differences between the streets and rivers conditions in Experiment 3? This difference must be conceptual, because the two conditions shared exactly the same stimulus items. Participants' decisions clearly depended on their knowledge of the categories (streets vs. rivers) that our cover story supplied. This suggests, in turn, that the sortal theory might have been right after all in asserting that basic-level concepts play a crucial role in defining individual identity. A more general way of putting this problem is to ask how the model weights different types of causal properties or kinds of causal continuity. Doesn't the model have to take into account aspects of the domain other than causal factors in fixing these weights?

In considering the idea of domain specificity, however, it is important to distinguish the source of the differences between

domains. All theories of identity must acknowledge that objects of different types can vary in their behavior in ways that are important for identity and persistence. Dropping a wine glass on a slate floor from a height of 3 feet will probably cause it to shatter and go out of existence, whereas dropping a cat on the same floor from the same height will probably leave it unscathed. However, this kind of domain specificity does not distinguish between the sortal and the causal continuer approaches. What does distinguish the theories is the explanation for such differences. In the case of the sortal view, the source of such differences is the meaning of the sortal terms that describe the objects. Part of the meaning of *glass*, for example, is an identity condition that stipulates that nothing after a shattering event can be identical to the original glass. By contrast, the causal continuer theory accounts for the difference in terms of the kinds of causes responsible for maintaining the integrity of the object in question. It is an empirical fact, and not part of the meaning of *glass* or *cat*, that some of the causes that disrupt a glass's existence do no damage to a cat. Similarly, participants' responses in the streets and rivers conditions of Experiment 3 may vary because of differences in causal (or other) information they have about these domains; however, this fails to show that the differences derive from the meanings of *street* and *river*.

An analogy may make this distinction clearer. The internal temperature of objects varies by domain; some types of objects have systematically higher temperature than others. The body temperature of birds, for example, tends to be higher than that of humans under normal conditions. In a sense, then, body temperature could be said to be domain specific. However, no one would suppose that the meanings of the terms *bird* and *human* include temperature conditions that specify the allowable range of body temperatures in these species. Instead, the temperature of different creatures is the result of mechanisms of thermal regulation, ambient temperatures, and other causal factors. In a parallel way, the causal continuer theory claims that domain differences in identity are due to differences in the kinds of causal mechanisms that maintain an object during its career rather than to differences in the meaning of expressions for these objects.

We are arguing that, in people's intuitive view, the meanings of sortals do not confer identity anymore than they confer temperature on otherwise temperature-less objects. Nevertheless, it is possible for a critic to contend that sortals still have an essential role to play in detecting identity. According to this modified sortal view, unless we know what sort of object we are dealing with, it is impossible to determine whether the physical features it exhibits are enough to establish its identity with another object. Unlike the case of temperature, in which we have thermometers that measure temperature in a truly domain-independent way, we do not have an internal identity detector that can operate without knowledge of the domain in question. This modified viewpoint could concede that causal factors are important elements in such identity judgments, but it could still maintain that what makes for causal integration in the case of cats differs from that of wine glasses in a way that is impossible to detect without additional principles concerning cats and glasses. The issue here is a delicate one of conceptual priority: Can we build up sufficient identity information about different types of objects solely through knowledge of their causal interactions, or do we have to appeal to some further conceptual principles about the domain?

Perhaps the best case for a domain-specific view comes from objects that are composed of materials that can exist independently, such as a statue and the lump of clay that makes it up (see Lowe, 2002, for a review of approaches to problems of composition). (It may be possible to question whether a lump of clay is an object in its own right, but let us assume for the sake of the argument that this is so.) If the lump of clay is reformed into a ball, then it would seem that the statue has ceased to exist while the lump of clay continues. However, because the very same causal forces exert their effects on both the statue and the lump during the time they coexist, does not this prove that causal forces are not sufficient to determine identity? It seems at first glance that, to trace the identity of the appropriate object, we need to know whether we are talking about the statue or the lump.

However, even composition cases fail to clinch the case for domain specificity (in the sense of principles associated with specific domains that are not attributable to causal differences between these domains). Although the same causes act on the lump and statue while they are merged, the causal stream responsible for the persistence of the lump may be distinct from that responsible for the statue. These streams separate when the lump is reshaped, at which point the causal factors responsible for the statue's existence—static forces that produce the statue's form—disperse. Similarly, if the statue is repaired with other materials, the causal stream underlying the statue is maintained, but not the stream upholding the lump.

#### *How Does Category Information Contribute to Object Identity?*

Reflection and thought experiments may be unable to decide whether there is some further principle associated with sortal or basic-level categories that we need in determining object identity. What about real experiments? Experiment 3 did not attempt to distinguish the causal continuer and sortal views, but Experiment 1 did, with results that favored causal continuity. Are there countervailing empirical reasons to favor a sortal approach? Developmental research using Xu and Carey's (1996) "Is-it-one-or-two?" paradigm suggests that even those infants who are able to succeed at the task when the transformation crosses basic-level boundaries fail when the transformation is within these boundaries (Bonatti et al., 2002; Xu et al., 2004). In Xu et al. (2004, Experiment 4), for example, infants habituated to a scene in which two distinct objects appeared in alternation at the right or left of an occluding screen. In one condition, however, the two objects were from different basic-level categories (e.g., a cup vs. a ball), whereas in a second condition the objects were from the same basic-level category (e.g., two cups) with similar size and color but different shapes (e.g., one cup with a top and two handles vs. one with no top and one handle). The results from the same-category condition showed that infants looked longer when the screen was removed to reveal two objects than one. This was approximately the same pattern they exhibited in a baseline condition in which the infants were exposed to the objects without prior habituation. In the different-category condition, however, they looked about the same amount of time at one as at two objects. Xu et al. (2004) interpret this result to mean that the infants were able to use information about the different categories to infer that there were two objects during habituation. When the screen was removed to reveal only one



object, they found the result surprising enough to eliminate their baseline preference for looking at two items. They made no such inference in the same-category condition, despite the objects' shape change, and hence looked just as long as in the baseline trials.

An initial question about these results is whether the shape change in the same-category condition (e.g., between two types of cups) was comparable to that of the different-category condition (between a cup and a ball). As Xu et al. (2004, pp. 180–181) discuss, the evidence on this point is mixed. However, putting this issue to the side, experiments of this type cannot distinguish the sortal from the causal continuer hypothesis, because basic-level category differences are confounded with causal differences. The type of transformation necessary to change one kind of cup into another may be causally less remote than that needed to change a cup into a ball. To unconfound sortal and causal factors, it is necessary to break the everyday correlation between them, as in the stories we used in Experiment 1 (and in the similar hypothetical scenarios of Blok et al., 2005). Obviously, hypothetical or counterfactual scenarios have their own methodological disadvantages. However, it is difficult to escape the need for these cases if the goal is to provide a clean test of the theories.

As we noted earlier, there is no reason to deny that sortal or basic-level categories can play a role in identity judgments. Experiment 1 showed that, contrary to Principle 3, membership in contrasting sortals does not entail nonidentity. Nevertheless, membership in contrasting sortals is presumptive evidence for nonidentity. Because hedgehogs tend not to turn into croquet balls, we can use the fact that one object is a hedgehog and another a croquet ball to decide they are different objects. However, this very plausible heuristic use of category information does not establish that sortals are conceptually prior to causal continuity (in the sense that we discussed in the previous section) or that “we must be supplied a sortal to trace identity” (Carey, 1995, p. 108). If category information plays merely a heuristic role, then infants' failure to use this information in the “Is-it-one-or-two?” task does not necessarily show that they lack the relevant basic-level concepts. That is, it is no longer straightforward to infer from their failure in this task that “very young infants have not yet constructed concepts that serve as adultlike meanings of words like ‘bottle,’ ‘ball,’ and ‘dog’” (Carey, 1995, p. 128). Just as an adult who lacks knowledge of how cars work could still have the concept of *car*, so a 10-month-old who lacks the heuristic knowledge that cups do not turn into balls could still have the concept of *cup*.

It is also worth noting that there are many methodological differences between studies that tend to support sortal theories (e.g., Bonatti et al., 2002; Xu & Carey, 1996; Xu et al., 2004) and those that suggest a more limited role for sortals (e.g., Blok et al., 2005; Liittschwager, 1995; and the present experiments). The former have used the “Is-it-one-or-two?” task paradigm, visually present objects, looking time measures, and infant participants; the latter have used stories about object transformations, explicit identity judgments, and older children or adults. These differences leave open many questions about the nature of developmental changes in the information and procedures people use in deciding on object sameness.

### *Transitivity of Identity*

Although the causal continuer model provides a good quantitative account of the data from our first two experiments, this accomplishment depends on its liberal policy with respect to “both” responses. The model produces these responses when both items are causally close enough to be the original but when the difference between the possible continuers is small enough to be ignored. This assumption is consistent with participants' responses: It is striking that in Experiment 1, when both contenders consisted only of particles copied from the original, nearly all participants made a “both” response (see Figure 3, bottom right). Similarly, participants in Experiment 2 often produced “both” responses when the two successor icebergs were attractive options (see, e.g., the case of two  $1 \times 1 \times 1$ -m icebergs in the upper left-hand graph in Figure 5). The trouble is that these responses appear to violate the transitivity property of identity in Principle 1c. How could both copies be identical to the original while not being identical to each other?

It is possible to view these responses as errors in participants' thinking about identity. Perhaps participants' identity decisions reflect a simple heuristic rather than a studied, normatively appropriate procedure. On this approach, the responses are much like intransitivities in the preference judgments of individual decision makers (Tversky, 1969). For example, participants in Experiments 1 and 2 may have used the causal distance between the original item and the copies, without concerning themselves with the extra constraints that identity imposes.<sup>16</sup> The tendency to choose “both” may be strengthened by the fact that, just after an object fissions, there may be little to distinguish the two continuers. Although there were two numerically distinct animals in the two-copy condition of Experiment 1, these animals presumably look alike, think the same thoughts, have the same preferences, and so on. If we were to ask about the identity of the same pair when they had developed their own distinctive properties, participants may reason that causal forces had driven the individuals apart, and they may be less likely to produce a “both” response. Moreover, when both alternatives are causally close, the four-alternative forced choice may not have included an option that fit participants' assessment of the situation. Perhaps they thought that one or the other item must be the original but that there was simply no way to decide between them. On this view, if we had included an option of the sort “it's one or the other but I can't tell which,” we would have reduced the number of “both” answers.

<sup>16</sup> Perhaps the tendency to make “both” responses was strengthened in Experiment 1 by the fact that the instructions referred to the successor objects as “copies.” This may have led participants to believe that neither item was the real thing and to respond on the basis of qualitative similarity rather than numerical identity. It is unclear, however, why participants who believed that the copies could not be the original would not have produced “neither” responses rather than “both” to our question about which is Fred. In addition, “both” responses also showed up in Experiment 2, where we made no reference to copies. It is possible, of course, that different factors were operative in the two experiments, but it is worthwhile pursuing the idea that a single explanation applies to both.

Alternative accounts of the “both” responses explain them as coherent ways of interpreting the experimental situation that avoid intransitivity. We have been assuming that participants believe the two copies in our experiment are distinct individuals, and this assumption leads to intransitivity when both copies are also identical to the original. Another way to view the situation is that splitting produces not two independent objects but two parts of a single temporally branching one (this is one of the individuals or “lifetimes” that Perry, 1972, discerns in fission cases). Figure 8a schematically illustrates this approach. The arrows in the diagram indicate the temporal sequence of events in the life of the lion, from its birth at  $t_0$  to the point at which it is copied at  $t_1$  to its end state at  $t_2$ . The boundary around these stages groups the stages that are the parts of the life of a single lion. According to this way of thinking, the duplicated lion in the stimulus stories exists after division in something like the way that a tree exists spatially in multiple branches in going from bottom to top. Just as the branches are parts of the same tree, the multiple copies are parts of the same creature. No intransitivity appears on this interpretation: Copy 1, Copy 2, and the original object are all the same individual.

A second way of salvaging transitivity is to construe the two copies as distinct objects but ones that existed all along, sharing the spatial parts of the original (Lewis, 1983). Figure 8b illus-

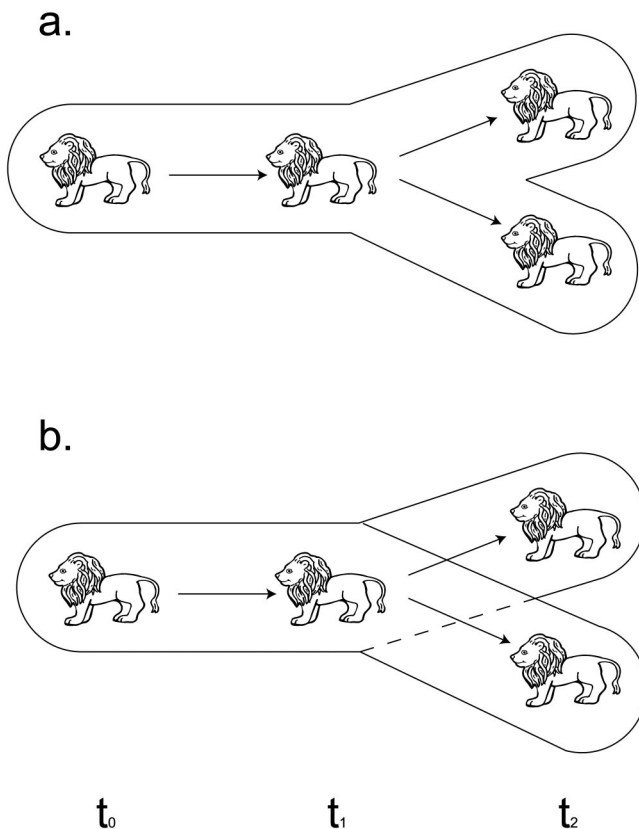


Figure 8. Two ways of interpreting fission examples. a: The original and the copies are temporal and spatial parts of a single branching object. b: Both copies are distinct objects that overlap spatially during the initial stages of their lives and diverge thereafter.

trates this reinterpretation. Copy 1 begins life when the original does, surviving the division and continuing on its own way. Copy 2 does the same. What is unusual about these individuals is that they are indistinguishable during the pre-fission part of their existence: What seemed to be a single original object turns out to be two cohabitators. Intransitivities also disappear on this interpretation, as they did with branching objects: When participants say that Copy 1 is still Fred, they mean that he is still Fred<sub>1</sub>, one of the two coembodied creatures, and when they say that Copy 2 is still Fred, they mean he is still Fred<sub>2</sub>, the other coembodied lion. Copy 1 = Fred<sub>1</sub>, Copy 2 = Fred<sub>2</sub>, but because these two are distinct, the judgments do not violate transitivity.

To understand these explanations of the “both” judgments, we conducted a pilot study based on scenarios similar to those of Experiment 2. Participants read about an iceberg, Sample 94, that splits into two equal-sized halves: one located to the east and the other to the west. On separate trials, participants learned that the resulting halves had either identical or different features (square vs. circular depressions). In one condition, the participants decided whether, after the split, only the eastern-most iceberg, only the western-most iceberg, both, or neither was Sample 94. They then decided how many different icebergs in total had existed during the entire time period from just before to after the split: one, two, or three. Finally, they were asked to justify their decisions to the questions. In a second condition, participants made the same judgments, except that instead of the four-choice decision they made a five-choice decision that included the additional option that “either the eastern-most or the western-most iceberg [is Sample 94] but one can’t tell which.”

We expected that if the “both” responses were due to carelessness, then asking participants for justifications would reduce “both” choices to low levels. Similarly, if the “both” responses were due to participants being unable to say that one of the icebergs was Sample 94 but that they were not sure which one, then providing this option would also reduce or eliminate “boths.” Neither of these predictions, however, was correct. Participants responded “both” on 59% of the trials when they had four choices and 51% when they had five choices. “Both” was, in fact, the most common response; nearly all the remaining judgments were “neither” (40% in the four-choice and 46% in the five-choice condition). Only 3 of the 20 participants in the five-choice condition made use of the response that Sample 94 was “either one or the other but one can’t tell which.” Likewise, specifying different features for the descendant icebergs did not greatly change the number of “both” responses (59% of responses were “both” when the icebergs had the same features and 50% when they had different features).

More interesting results come from a comparison between answers to the questions “Which is Sample 94?” and “How many icebergs existed?” As might be expected, participants who decided that neither was Sample 94 usually said there were three objects involved: Sample 94 and the two icebergs resulting from the split. (Participants said three icebergs existed on 85% of “neither” trials.) “Both” responders, however, produced a more even distribution of answers to the “How many?” question: Participants responded that one object existed on 34% of both trials, two objects on 18%, and three objects on 48%. Although the justifications were not always informative, some provided insight into participants’ reasoning and, in fact, provided some limited support

for the alternative explanations just discussed. “Both” responders who said that only one object existed tended to appeal to the idea that splitting did not produce anything new, perhaps the two resulting chunks of ice are temporal branches of the original, or the two resulting chunks of ice are simply not full-fledged icebergs in their own right (“the same just in different locations,” “[Sample] 94 was determined before the split, so both are [Sample] 94,” “the ice that the new ones are composed of isn’t new ice,” “just like a cookie breaking into two half-moon shapes, still one cookie”).

The critical responses come from participants who believed that both resulting icebergs were Sample 94 but that two or three icebergs existed altogether. These responses are the clearest indicators of genuine inconsistencies, because if both of the resulting objects are Sample 94, how could there be two or three distinct objects? It is difficult to interpret the justifications from “both” responders who said that two icebergs existed, although in a few cases participants may have believed two icebergs were coembodied in the original item (“Sample [94] was really 2 icebergs,” “two still [Sample] 94, E[ast] and W[est]”), as in Lewis’s (1983) theory. “Both” responders who believed that three objects existed tended to emphasize in their justification the presence of three physically separated objects but then also cited unifying factors, such as the part-whole relation (“both parts; two new, plus original,” “both parts of Sample 94; 1 original plus the 2 = 3,” “if object = sum of its parts, both = 94; however, still 3 exist: original and offspring”) or the presence of only a single name (“[descendants] were not renamed, so same name as original; 1 and 2 new = 3,” “no contrary evidence of new labels, so continue to be 94,” “samples not renamed,” “one, no renaming; but physically 3 objects”).

Although these results are preliminary, they suggest that “both” answers are not due to simple response factors (e.g., absence of a “don’t know which” alternative). Instead, they seem to reflect participants’ ways of reconciling potentially conflicting information in the scenarios, including physical segregation, part-whole relations, lack of new material substance, and lack of new names. Except for the last of these factors, these relations are inherent in the causal scenarios we described. In case the unifying factors (e.g., part-whole relations) prevail so that participants see the situation as containing a single object, “both” responses do not indicate true intransitivities. When unifying factors do not prevail, however, “both” responses may reflect intransitivities because of participants’ inability to resolve the conflicting factors (e.g., spatial segregation vs. part-whole status). The causal continuer model does not force a resolution on the participants and thus predicts the “both” responses. This seems the right tact for a descriptive cognitive theory that must deal with this outcome. However, further elaboration of the model would be necessary to explain the fine-grained reasoning (as in the justifications we have quoted) that leads people to say “both.”

### Summary and Conclusions

The causal continuer approach contends that a later manifestation of a single object must causally grow from earlier ones, so that causality takes precedence over qualitative overlap in properties, spatiotemporal continuity, or sortal membership. Similarity, continuity, and other properties can come into play, however, if direct causal information is absent or ambiguous. The model makes its

identity judgments on the basis of two interrelated decisions: An object  $x_0$  is identical to another  $x_1$  if  $x_1$  is causally close enough to be the continuation of  $x_0$  and if  $x_1$  is the closest of all the close-enough competitors. This is the answer the theory gives to the central issue, Question 1, that we started with.

Evidence for this approach comes from studies in which we manipulated the closeness of an original object to each of two possible continuers. In the first of these experiments, the model succeeded in predicting participants’ decisions about which continuer was identical to an original object in a setting in which there was no spatial continuity between the items. The same study showed a dissociation between these identity judgments and judgments of basic-level (or sortal) category membership. Experiment 2 produced evidence for a similar dissociation between identity and similarity, and it showed that the model could also predict identity judgments in cases of splitting or shrinking (and when information about spatial continuity was not available). A third study extended the model’s predictions to a domain in which competing causal continuers (branches of streets and rivers) are more ordinary entities. The model’s two subparts predicted corresponding differences when participants decided whether either branch had the same name as the original.

The causal continuer approach seems capable of handling many of the issues that created stumbling blocks for earlier theories. Because the model subordinates similarity judgments to causal continuity, it explains why similarity can function as evidence for identity in some situations but as evidence against it in others. For example, a difference in size (a dissimilarity) may support the hypothesis that the iceberg you perceive now is the same one you saw earlier but may contradict the hypothesis that the cup you perceive now is the same as an earlier one. Experiment 2 showed that judgments of similarity and identity diverge in cases in which people expect change over time. Moreover, by settling on causal connectedness as the basis for identity, the causal continuer approach also avoids problems of circularity associated with similarity.

Along the same lines, although knowledge of spatiotemporal continuity is an important clue to sameness, it need not be decisive. In the vicinity of Armstrong-Nozick-Shoemaker machines, for example, causal facts about the devices block the inference from the continuity of two items to the conclusion that the later one is a causal outgrowth of the earlier. We need not resort to any kind of spatiotemporal continuity if we already know the causal facts. For the same reason, the causal continuer theory is not vulnerable to counterexamples in which spatial coherence is disrupted (e.g., Hirsch’s, 1982, disassembled device). Causal forces can extend through the disassembled pieces of an object, rejoining in the reassembled object. We have seen, in fact, that the causal continuer theory can deliver correct predictions in cases of radical disassembly in Experiment 1.

Finally, because ordinary objects rarely switch between basic-level categories, knowledge of category membership can be a clue to object identity. If you see a table and then a chair, it is probably safe to conclude that they are different individuals. In such cases, however, the category distinction is correlated with a variation in causal distance. It is easier for causal factors to produce from the original table appearance a second table appearance than a chair appearance. When Experiment 1 broke this correlation, participants’ judgments followed the causal pathway.

Our concepts of people and other things must be rich enough to support conjectures about what might have happened to these individuals in situations that are possible but never actually take place. Concepts of our friends, for example, inform our guesses about how they will behave in settings they have not yet, and perhaps never will, experience. The same goes for predictions about political figures or celebrities whose dispositions we think we know. A trend in historical writing exploits such concepts by considering counterfactual situations in which these individuals figure (e.g., McNeill, 2001; Rabb, 2001; Tally, 2000). In the realm of inanimate objects, predictions about location and change have the warrant of well-established physical principles, even when the predictions' initial conditions never occur. According to the causal continuer theory, what give us the ability to make these counterfactual judgments are the same causal relations that govern our ability to trace these individuals in the real world.

### References

- Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts? *Cognition*, *69*, 135–178.
- Anisfeld, M., & Knapp, M. (1968). Association, synonymy, and directionality in false recognition. *Journal of Experimental Psychology*, *77*, 171–179.
- Armstrong, D. M. (1980). Identity through time. In P. van Inwagen (Ed.), *Time and cause* (pp. 67–78). Dordrecht, The Netherlands: Reidel.
- Ayers, M. (1997). Is physical object a sortal? A reply to Xu. *Mind & Language*, *12*, 393–405.
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Bjork, R. A. (1978). The updating of human memory. *Psychology of Learning and Motivation*, *12*, 235–259.
- Blok, S., Newman, G., & Rips, L. J. (2005). Individuals and their concepts. In W.-K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in Honor of Douglas L. Medin* (pp. 127–149). Washington, DC: American Psychological Association.
- Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, *44*, 388–426.
- Bower, G. H., & Glass, A. L. (1976). Structural units and the reintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 456–466.
- Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*, 14–21.
- Carey, S. (1985). *Conceptual change in childhood*. 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.), *Invitation to cognitive science: Vol. 3. Thinking* (2nd ed., pp. 101–129). Cambridge, MA: MIT Press.
- Carey, S., & Xu, F. (1999). Sortals and kinds. In R. Jackendoff, P. Bloom, & K. Wynn (Eds.), *Language, logic, and concepts* (pp. 311–335). Cambridge, MA: MIT Press.
- Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, *80*, 179–213.
- Chisholm, R. M. (1976). *Person and object*. LaSalle, IL: Open Court.
- Conway, M. A. (1990). *Autobiographical memory*. Milton Keynes, UK: Open University Press.
- Davis, N. Z. (1983). *The return of Martin Guerre*. Cambridge, MA: Harvard University Press.
- Dummett, M. (1973). *Frege: Philosophy of language*. Cambridge, MA: Harvard University Press.
- Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Oxford, UK: Blackwell.
- Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 234–257.
- Gelman, S. A. (2003). *The essential child*. Oxford, UK: Oxford University Press.
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the nonobvious. *Cognition*, *38*, 213–244.
- Gutheil, G., Bloom, P., Valderrama, N., & Freedman, R. (2004). The role of historical intuitions in children's and adults' naming of artifacts. *Cognition*, *91*, 23–42.
- 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.
- Hall, D. G., Lee, S. C., & Bélanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology*, *37*, 298–307.
- Hall, D. G., Waxman, S. R., Brédart, S., & Nicolay, A.-C. (2003). Preschoolers' use of form class cues to learn descriptive proper names. *Child Development*, *74*, 1547–1560.
- Hirsch, E. (1982). *The concept of identity*. Oxford, UK: Oxford University Press.
- 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. (Original work published 1655)
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.
- 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.
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*, 141–221.
- Kihlstrom, J. F., & Cantor, N. (1984). Mental representations of the self. *Advances in Experimental Social Psychology*, *17*, 1–47.
- Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology*, *63*, 739–753.
- Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., & Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measures. *Biometrics*, *33*, 133–158.
- Kolak, D., & Martin, R. (1987). Personal identity and causality: Becoming unglued. *American Philosophical Quarterly*, *24*, 339–347.
- Kripke, S. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5 month old infants see humans as material objects? *Cognition*, *94*, 95–103.
- Lewis, D. (1983). Survival and identity. In *Philosophical papers* (Vol. 1, pp. 55–77). Oxford, UK: Oxford University Press.
- Liittschwager, J. C. (1995). Children's reasoning about identity across

- transformations. *Dissertation Abstracts International*, 55(10), 4623B. (UMI No. 9508399).
- Lowe, E. J. (2002). *A survey of metaphysics*. Oxford, UK: Oxford University Press.
- Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41, 954–969.
- McNeill, W. H. (2001). What if Pizarro had not found potatoes in Peru? In R. Cowley (Ed.), *What if? 2* (pp. 415–427). New York: Berkley Books.
- 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, UK: Cambridge University Press.
- Medin, D. L., & Rips, L. J. (2005). Concepts and categories: Meaning, memory, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 37–72). Cambridge, UK: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Mendelson, E. (1964). *Introduction to mathematical logic*. New York: Van Nostrand.
- Millikan, R. G. (2000). *On clear and confused ideas*. Cambridge, UK: Cambridge University Press.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Noonan, H. W. (1985). The closest continuer theory of identity. *Inquiry*, 28, 195–229.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Oshima-Takane, Y. (1999). The learning of first and second person pronouns in English. In R. Jackendoff, P. Bloom, & K. Wynn (Eds.), *Language, logic, and concepts* (pp. 373–409). Cambridge, MA: MIT Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford, UK: Oxford University Press.
- Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology*, 51, 907–917.
- Perry, J. (1972). Can the self divide? *Journal of Philosophy*, 69, 463–488.
- Peterson, M. A. (2001). Object perception. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 168–203). Oxford, UK: Blackwell.
- Rabb, T. K. (2001). If Charles I had not left Whitehall, August 1641. In R. Cowley (Ed.), *What if? 2* (pp. 122–133). New York: Berkley Books.
- Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, 130, 323–360.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, UK: Cambridge University Press.
- Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin*, 127, 827–852.
- Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As time goes by: Children's understanding of growth in animals. *Child Development*, 62, 1302–1320.
- Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ: Erlbaum.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49, 11–36.
- Shoemaker, S. (1979). Identity, properties, and causality. *Midwest Studies in Philosophy*, 4, 321–342.
- Sider, T. (2001). *Four dimensionalism*. Oxford, UK: Oxford University Press.
- Sloman, S. A., & Malt, B. (2003). Artifacts are not ascribed essences, nor are they treated as belonging to kinds. *Language and Cognitive Processes*, 18, 563–582.
- Sloman, S. A., & Rips, L. J. (1998). Similarity as an explanatory construct. *Cognition*, 65, 87–101.
- Sorrentino, C. M. (2001). Children and adults represent proper names as referring to unique individuals. *Developmental Science*, 4, 399–407.
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56.
- Spelke, E. S., Gutheil, G., & Van de Walle, G. (1995). The development of object perception. In D. N. Osherson (Series Ed.) & S. M. Kosslyn & D. N. Osherson (Vol. Eds.), *Invitation to cognitive science: Vol. 2. Visual cognition* (2nd ed., pp. 297–330). Cambridge, MA: MIT Press.
- Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatio-temporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, 13, 113–142.
- Sternberg, R. J. (1982). Natural, unnatural, and supernatural concepts. *Cognitive Psychology*, 14, 451–488.
- Sternberg, R. J., Chawarski, M. C., & Allbritton, D. W. (1998). If you changed your name and appearance to those of Elvis Presley, who would you be? Historical features in categorization. *American Journal of Psychology*, 111, 327–351.
- Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research*, 38, 947–951.
- Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. London: Routledge.
- Stevens, M. (2000). The essentialist aspect of naive theories. *Cognition*, 74, 149–175.
- Tally, S. (2000). *Almost America*. New York: HarperCollins.
- Tversky, A. (1969). The intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Walker, J. H. (1975). Real-world variability, reasonableness judgments, and memory representations for concepts. *Journal of Verbal Learning and Verbal Behavior*, 14, 241–252.
- Wiggins, D. (1980). *Sameness and substance*. Cambridge, MA: Harvard Press.
- Wiggins, D. (1997). Sortal concepts: A reply to Xu. *Mind & Language*, 12, 413–421.
- Wiggins, D. (2001). *Sameness and substance renewed*. Cambridge, UK: Cambridge University Press.
- Wilcox, T., & Baillargeon, R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology*, 37, 97–155.
- Williams, B. (1982). Cosmic philosopher. *New York Review of Books*, 29, 32–34.
- 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. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85, 223–250.
- Xu, F. (2003). The development of object individuation in infancy. In H. Hayne & J. Fagen (Eds.), *Progress in infancy research* (Vol. 3, pp. 159–192). Mahwah, NJ: Erlbaum.
- Xu, F. (2005). Categories, kinds, and object individuation in infancy. In L.

- Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories in developmental time* (pp. 63–89). Mahwah, NJ: Erlbaum.
- Xu, F., & Baker, A. (2005). Object individuation in 10-month-old infants using a simplified manual search method. *Journal of Cognition and Development, 6*, 307–323.
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*, 111–153.
- Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*, 155–190.
- Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology, 76*, 495–523.

Received July 3, 2004

Revision received June 7, 2005

Accepted June 8, 2005 ■