Cognitive Psychology 65 (2012) 457–485

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Functions in biological kind classification

Tania Lombrozo [a,*], Bob Rehder [b]

[a] Department of Psychology, 3210 Tolman Hall, University of California, Berkeley, CA 94720, United States
[b] Department of Psychology, New York University, United States

ARTICLE INFO

Article history:
Accepted 20 June 2012
Available online 9 August 2012

Keywords:
Categorization
Functional features
Diagnosticity
Functions
Biological kind concepts

ABSTRACT

Biological traits that serve functions, such as a zebra's coloration (for camouflage) or a kangaroo's tail (for balance), seem to have a special role in conceptual representations for biological kinds. In five experiments, we investigate whether and why functional features are privileged in biological kind classification. Experiment 1 experimentally manipulates whether a feature serves a function and finds that functional features are judged more diagnostic of category membership as well as more likely to have a deep evolutionary history, be frequent in the current population, and persist in future populations. Experiments 2–5 reveal that these inferences about history, frequency, and persistence account for nearly all the effect of function on classification. We conclude that functional features are privileged because their relationship with the kind is viewed as stable over time and thus as especially well suited for establishing category membership, with implications for theories of classification and folk biological understanding.

1. Introduction

While traveling in Australia in 1770, the naturalist Sir Joseph Banks encountered an animal that shared several characteristics with the opossums he had observed in the Americas. In particular, the animal had a pouch with the function of holding young offspring. Largely on the basis of this resemblance, Banks identified the new find as "an animal of the Opossum tribe" (Wightman, 2008), leading to its current designation as a "possum." Banks' classification was unfortunate, however, as American opossums and Australian possums, though both marsupials, are only distantly related. Nonetheless, the mistake is understandable: Traits with biological functions, such as the opossum's

* Corresponding author.
E-mail address: lombrozo@berkeley.edu (T. Lombrozo).

0010-0285/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cogpsych.2012.06.002

pouch, seem to have a special relationship to category membership. For example, consider the close connection between the kangaroo and its tail (for balance), the zebra and its stripes (for camouflage), or the honeybee and its sting (for protection). In this paper we investigate whether and why functional features – those features that are viewed as promoting a species' continued survival – have a special status in classifying biological kinds.

A variety of researchers have suggested that functions are deeply embedded in folk-biological reasoning (Atran, 1994, 1995; Keil, 1992, 1994; Medin & Atran, 1999). Both children and adults prefer to explain biological traits by appeal to functions or purpose (e.g., Kelemen, 1999; Lombrozo & Carey, 2006; Lombrozo, Kelemen, & Zaitchik, 2007), even when doing so is potentially unwarranted (e.g., Gould & Lewontin, 1979; Kelemen & Rosset, 2009). For example, given the choice between explaining eyes by appeal to their role in seeing or by citing the cells that produce them, the majority of children and adults prefer the former (Kelemen, 1999; Lombrozo et al., 2007). The evolutionary biologist Richard Dawkins has gone so far as to suggest that when it comes to reasoning about the biological world, we humans "have purpose on the brain" (Dawkins, 1995).

Previous research also suggests that both children and adults privilege functional features when classifying artifacts, objects that are the product of intentional human design. For example, young children extend category labels from one object to another on the basis of a shared function (e.g., Kemler-Nelson & et al., 1995; see Oakes & Madole, 2008, for review), and adults will make categorization judgments on the basis of features relevant for an object's function over features that are equally or even more frequent in the relevant category, and thus more diagnostic of category membership (Lin & Murphy, 1997; Murphy & Wisniewski, 1989; Wisniewski, 1995). In fact, some studies suggest that having a particular function is necessary and sufficient for establishing an artifact's category membership (Barton & Komatsu, 1989; Keil, 1989), although other findings challenge the generality of such results and suggest that the relationship between functions and category membership is more complex (Ahn, 1998; Malt & Johnson, 1992). For example, some findings suggest an important role for the known or inferred intention of the artifact's creator (e.g., Bloom, 1996, 1998; cf. Chaigneau, Barsalou, & Sloman, 2004).

Despite these compelling indications that functions are central to biological reasoning and accorded a special status in classifying artifacts, research concerning the role of functional features in biological kind classification has been relatively sparse, and if anything challenges the conclusion that functional features are particularly important. DiYanni and Kelemen (2005) found that a majority of children judged that an artifact or living thing unable to perform a function (e.g., a lion that could no longer "run around") needed to be repaired or replaced by another category member, suggesting that the functional failure threatened the individual's category membership. However, adults showed a different pattern, judging that artifacts that no longer served a function needed to be repaired or replaced, but not living things. Barton and Komatsu (1989) found a similar domain difference with adults, with functional features more important than "compositional" (e.g., chromosomal or molecular) features for judging the category membership of artifacts (e.g., for a tire, the ability to roll was more important than being made of rubber), but less important for natural kinds (e.g., for a female goat, giving milk was less important for category membership than its chromosomal structure).

Finally, Ahn (1998) has proposed that any special influence that functional features have on biological classification is not due to functionality per se, but rather occurs only when functional features are thought to be the causes of other features (Ahn, Kim, Lassaline, & Dennis, 2000). Consistent with this proposal, Ahn showed that Barton and Komatsu's functional features were typically causes of the compositional features for artifacts but effects of the compositional features for natural kinds. Moreover, causes dominated classification decisions when Ahn experimentally manipulated whether functional features caused the compositional features or vice versa. However, more recent research has shown that the advantage for causal features does not always obtain and, when it does, it operates by lowering the importance of the effect feature rather than raising that of the cause (Rehder & Kim, 2010). In addition, Lombrozo (2009) has shown that a functional feature's special influence on categorization can obtain above and beyond its causal role in some conditions (e.g., when it is referenced in functional explanations for other features).

In sum, research on the status of functional features in biological kind classification has yielded mixed results, with some suggestions that functional features have a privileged status (especially for children), and others challenging the idea that functional features are in any way distinguished by virtue of serving a function. These results need to be interpreted with caution, however, as most research has focused on differences across development or across domains, and has not manipulated whether the very same feature does or does not have a function, thereby confounding functionality with other dimensions along which features can vary. Perhaps more importantly, there have been few explicit proposals concerning whether and why functional features might be privileged in biological kind classification, making it difficult to assess the implications of extant findings. These are important omissions given the ubiquity of functional features in biological kinds and the opportunity they afford to explore both theories of categorization and of folk biological understanding.

In the remainder of the introduction we present four hypotheses concerning why functional features might be privileged in biological kind classification. We then provide an overview of five experiments that aim to differentiate these hypotheses.

1.1. The affordance hypothesis

One possibility is that the role of functional features follows directly from the conceptual or semantic entailments of having a function – that is, from what it means for a feature to be functional. A small body of empirical research aims to characterize children and adults' concept of function (e.g., Barsalou, Sloman, & Chaigneau, 2005; Oakes & Madole, 2008), with a much larger body of work in philosophy (e.g., Allen, 2009). Broadly speaking, there have been two approaches to defining functions: ahistorical and historical. The first hypothesis we consider stems from the ahistorical approach, according to which a feature has a particular function if it causally contributes to a particular consequence (see Cummins, 1975, for a more sophisticated version of this account from philosophy). Thus the heart has the function of pumping blood because it causally contributes to blood pumping, but by the same logic it also has the function of making a thumping sound. Within psychology, this view has typically been characterized in terms of an object's affordances, appearance, or structural properties (e.g., Kemler-Nelson, Frankenfield, Morris, & Blair, 2000; Kemler-Nelson, Russell, Duke, & Jones, 2000). Some findings suggest that young children operate with an ahistorical or affordance-based understanding of function (e.g., Defeyter & German, 2003; Truxaw, Krasnow, Woods, & German, 2006; see Oakes & Madole, 2008 for review), and adults often use the word "function" in this sense, as when stating that noses serve the function of holding up glasses.

Features that are functional in this ahistorical sense could be central to classification because they indicate how category members and their features can be used. For example, privileging a biological trait's potential uses could help predict an organism's behavior or viability in a given environment. In what follows, we refer to the idea that functional features are important by virtue of their affordances *the affordance hypothesis* (see Table 1 for a summary of hypotheses with key claims and predictions).

1.2. The "functions are historical" hypothesis

In contrast to the ahistorical approach to function, the historical approach maintains that a feature has a particular function if it was created with the intention of serving that function or resulted from a process for which the function played a causal role in the creation, maintenance, or spread of that feature (Lombrozo & Carey, 2006; see Wright, 1973, in philosophy). For a biological organism, a feature is only functional in this sense if it resulted from a consequence-driven process, such as natural selection.[1] The heart, for example, has the function of pumping blood because the fact that hearts pump blood led to the maintenance and spread of hearts in biological populations. Hearts do not have the function of making a thumping sound because the thumping sound played no role in the history of hearts.

[1] Learning is also a consequence-driven process that can support a historical function. In the current paper, however, we focus on historical functions that resulted from phylogenetic processes such as natural selection, as our central hypotheses concern whether a feature has a deep (phylogenetic) versus a shallow (ontogenetic) history. Nonetheless, it is worth recognizing that the notion of a historical function – as advocated by some philosophers of biology – is logically orthogonal to whether a feature is phylogenetic or ontogenetic, as one can have an ontogenetic feature with a historical function (e.g., a learned behavior) or a phylogenetic feature without a historical function (e.g., a product of genetic drift that has an incidental affordance).

Table 1
Key hypotheses with corresponding claims and predictions.

Hypothesis	Claim	Predictions
Affordance hypothesis	Functional features are privileged in classification because of their potential causal contributions to outcomes (i.e., their affordances)	All else being equal, features with functions will be more diagnostic of category membership than nonfunctional features
Functions are historical hypothesis (*historical hypothesis*)	Functional features are privileged in classification because they are assumed to result from causal processes (such as evolution) that generated the category	All else being equal, features with a deep causal history will be judged more diagnostic, and effects of function should be eliminated (or reduced, if other hypotheses operate in concert) when causal history is controlled
Functions are frequent hypothesis (*frequency hypothesis*)	Functional features are privileged in classification because they are assumed to be frequent among (and therefore diagnostic of) current category members	All else being equal, features that are more frequent among current category members will be judged more diagnostic, and effects of function should be eliminated (or reduced, if other hypotheses operate in concert) when current frequency is controlled
Functions are stable hypothesis (*stability hypothesis*)	Functional features are privileged in classification because they are assumed to be frequent among category members (and therefore diagnostic) over time (i.e., past, present, and future)	All else being equal, features that are frequent among past, current, and future category members will be judged more diagnostic, and effects of function should be eliminated (or reduced, if other hypotheses operate in concert) when feature frequency in past, present, and future category members is controlled

Within psychology, historical approaches to function have tended to focus on the role of a designer's intentions in artifact categories, and find support in research on categorization and naming in both children and adults (e.g., Bloom, 1996, 1998; Casler & Kelemen, 2005, 2007; Diesendruck, Markson, & Bloom, 2003; Gelman & Bloom, 2000; German & Johnson, 2002; Matan & Carey, 2001; Puebla-Ramirez & Chaigneau, 2011; see Kelemen & Carey, 2007 for review). Researchers have also highlighted features' causal roles in the context of biological kind classification, but without the emphasis on their history. For example, features that appear earlier in a category's causal network of features tend to have greater weight in categorization decisions (Ahn et al., 2000; Sloman, Love, & Ahn, 1998). Numerous theorists have additionally noted that observable features of biological kinds can be used to infer the presence of invisible but essential features and thus category membership (Gelman, 2003; Hampton, Estes, & Simmons, 2007; Medin & Ortony, 1989; Murphy & Medin, 1985; Rehder, 2003, 2007; Rehder & Kim, 2009, 2010; but see Strevens, 2000). Inspired by historical approaches to function, we extend this inferential view of biological classification by suggesting that a feature's functionality for a kind implies that the feature arose from historical causal processes associated with the origin of the kind (e.g., natural selection), and that this lineage leads people to treat the feature as more diagnostic of kind membership. We call this hypothesis the *functions are historical* hypothesis (or the *historical* hypothesis for short).

1.3. The "functions are frequent" hypothesis

Where the first two hypotheses emphasize the conceptual beliefs associated with functional features, the third hypothesis considers how having a function influences people's beliefs about how widely a feature is distributed among category members. According to the *functions are frequent* hypothesis (or *frequency* hypothesis for short), functional features are assumed to be more frequent among category members, which in turn means that (all else being equal) they will be more diagnostic of category membership (e.g., Hampton, 1979; Nosofsky, 1988; Rosch & Mervis, 1975). Note that whereas frequency (also referred to in the categorization literature as *category validity*, i.e., the probability of the feature given the category) is usually construed to reflect classifiers' direct observation of category members, here we assume that it is an inference they make on the basis of a feature's function.

One motivation for the frequency hypothesis comes from the sorts of beliefs that people might have about the causal processes that generate functional features and govern their interactions with other features and the environment. Consistent with this idea, previous research has shown that causal beliefs can influence classifiers' subjective beliefs about a feature's frequency. For example, Rehder and Kim (2010) found that features were rated as more prevalent in category members and more important to category membership to the extent they had stronger causes (also see Sloman et al., 1998). Conceivably, the beliefs that people have about how functions causally interact with other features and aspects of the organism's environment might lead them to believe that useful (i.e., functional) features are more widespread than those that are not. In the general discussion we elaborate on the nature of the causal beliefs that might warrant this inference. For now, the key empirical prediction is that a feature's function serves as a cue to its frequency among category members, which in turn increases its diagnosticity when it comes to category membership.

1.4. The "functions are stable" hypothesis

Our final hypothesis is motivated by the observation that categories are useful not only for reasoning about the present, but also to describe and explain the past (e.g., Murphy & Medin, 1985; Prasada & Dillingham, 2006, 2009; Rips, 1989) and to anticipate potential future or counterfactual possibilities (e.g., Keil, 1989; Rips, 1989; Sloman et al., 1998; see also Markman & Ross, 2003). These considerations suggest that features are likely to be judged more important if they are assumed to be *stable,* that is, frequent (and thus diagnostic) not only in the present but also in the past and future. The *functions are stable* hypothesis (or *stability* hypothesis for short) maintains that functional features are especially likely to satisfy these criteria. Like the frequency hypothesis, the stability hypothesis is potentially grounded in people's beliefs about causal processes: The same processes responsible for a feature's present existence could imply its existence in the past, work to maintain its presence in the future, and even ensure that its existence is robust in the face of actual or counterfactual change. The stability hypothesis thus subsumes not only the frequency hypothesis (in that it explains why current frequency should matter), but possibly the historical hypothesis, since causal history could inform classifiers' assessment of whether a feature was prevalent in the past. Again, we develop these ideas further in the general discussion. For now, the key idea is that functional features are assumed to be frequent over time and so share a particularly stable relationship with category membership.

It is important to emphasize that the hypotheses specified above are not mutually exclusive, as nothing rules out the possibility that functional features are important because of their affordances and also because of the inferences they license about causal history and frequency. In fact, we have already mentioned how the stability hypothesis potentially subsumes the historical and frequency hypotheses. Our hypotheses are also not exhaustive, as there are yet other possibilities that have been entertained regarding the source of functional features' importance (e.g., that they tend to be the cause of other features, Ahn, 1998; enter into functional explanations, Lombrozo, 2009; instantiate an ideal associated with biological kinds, Barsalou, 1985; and so forth). We return to each of these possibilities in the general discussion.

1.5. Overview of experiments

We now present five experiments that investigate the hypotheses identified above. Our primary goal is not necessarily to identify a single "best" hypothesis, but rather to establish whether and why the factors identified by each hypothesis contribute to the relationship between function and classification.

Experiment 1 has two aims. The first is to investigate whether functional features are indeed privileged when it comes to classifying biological kinds. Because the functional features of natural categories can differ from the nonfunctional ones on many dimensions (e.g., familiarity, salience), we adopt an experimental approach in which participants are taught novel categories with features that either do or do not have functions. For example, participants learn about a novel biological kind named Sacramento Ants. Half the participants are simply told that Sacramento Ants are red, whereas the other half are additionally told that red provides camouflage. By comparing judgments across groups of participants, we examine whether features that serve a function are judged more diagnostic of category membership than features that do not.

The second aim of Experiment 1 is to break new ground in understanding why functional features influence classification by gathering preliminary evidence regarding the hypotheses presented above. We do so by examining the additional inferences that a feature's functionality licenses. First, we ask participants to rate whether features have a deep evolutionary history. A finding that functional features are rated more likely to have such a history than nonfunctional features would provide support for the historical hypothesis. Second, we ask participants to rate the prevalence of features in the current population of category members, as well as whether the features are likely to persist in future populations. A finding that functional features are viewed as more frequent in current category members would provide support for the frequency hypothesis, and the finding that they are also viewed as more likely to persist in the future would support the stability hypothesis.

To foreshadow the results of this initial experiment, we find that participants indeed view functional feature as more important to category membership. Importantly, they also rate such features as more likely to have a deep causal history and to be more prevalent in current and future category members, consistent with the possibility that these variables mediate the relationship between function and categorization. However, it is also possible that these variables *reflect* features' importance rather than being a *cause* of it. For example, features could be judged to have a deep causal history *because* they are judged important to categorization. Accordingly, the remaining experiments explicitly manipulate these additional variables in order to establish their causal influence on classification (see Table 1 for a summary of predictions). First, Experiment 2 investigates the role of causal history by orthogonally varying whether a novel biological kind's features support a function and have a deep causal history. Because the historical hypothesis claims that functional features' importance is mediated by their history, it predicts the absence (or reduction) of an effect of function on classification when history is controlled. Experiment 3 tests whether function's effect on classification is mediated by assumptions about the frequency of functional features amongst category members. The frequency hypothesis predicts the absence (or reduction) of an effect of function when a feature's current frequency is controlled. Finally, Experiments 4 and 5 focus on the stability hypothesis and test whether functional features are central for biological kinds not only because of the inferences they license about the feature in past or present populations, but also because of the inferences they license about the future.

2. Experiment 1: Functional features

Experiment 1 varied whether the features of a novel biological kind did or did not have a function. For example, one such kind was the Sacramento Ant, which typically has thick blood, a red color, sticky antennas, and slow digestion. Each feature was described as either functional or nonfunctional. After introducing each novel kind, we solicited judgments about that feature's importance for category membership by asking how likely it was that an organism *missing* that feature was a category member, and we additionally asked about the feature's evolutionary history, prevalence in the population, and persistence in future populations.

2.1. Methods

2.1.1. Participants

Eighty participants were recruited on-line from Amazon Mechanical Turk (59% women, mean age 33) and paid in exchange for their participation.[2] Using Mechanical Turk's qualification settings, partic-

[2] An additional 11 participants were excluded for failing an instructional manipulation check. At the beginning of the experiment, participants were presented with the following text: "You will learn about three kinds of plants or animals. Please read the information carefully and answer the questions that follow. To indicate that you have read the instructions, please select the second button, the one labeled "select to go back," to proceed." This was followed by two buttons prominently labeled, "Select to continue" and "select to go back." To pass the instructional manipulation check, participants had to select the second option. (See Oppenheimer, Meyvisb, & Davidenkoc, 2009, for similar technique.) However, all significant results remain significant when the 11 participants excluded on this basis are included in analyses.

ipation was restricted to people in the United States and who had an approval rate of 95% or higher (i.e., their "HIT approval rate" was greater than or equal to 95).

2.1.2. Materials

Three novel categories were tested: Sacramento Ants, Albany Ferns, and Rwandan Marmots. Four features were specified for each category, and each feature could appear in one of two versions: functional or nonfunctional. These materials were used to create two stimulus sets such that each category involved two functional and two non-functional features, and each feature was functional in one stimulus set and nonfunctional in the other. Sample features for Sacramento Ants are provided in Table 2.

2.1.3. Procedure

The study was administered on-line using web-based experiment presentation software. Participants were told that they would learn about three categories of plants or animals. For each category, they read feature descriptions for each of four features and then completed categorization judgments, historical inference judgments, frequency estimation judgments, and future inference judgments, as detailed below.

For the categorization task, participants were asked to imagine coming across members of the category's superordinate class that possessed all but one of the novel category's features. For example, for the Sacramento Ant category, participants were asked to imagine coming across four ants, where one had every feature typical of Sacramento Ants except thick blood, another every feature except a red color, and so on for each of the four Sacramento Ant features. For each such specimen, participants judged how likely it was that the specimen was a member of the novel category, and indicated their judgment on a 9-point scale ranging from "very unlikely" to "very likely," with the mid-point labeled "neither likely nor unlikely." The four specimens evaluated were presented in a random order.

For the historical inference judgments, participants were presented with the following prompt: "Some features are primarily the result of evolutionary history, while others are the result of experiences during an organism's recent history (e.g., things that may have changed in the species' environment). For each feature below, rate how likely you think it is that the feature results from evolutionary processes." This was followed by all four features for a given category in random order, with a 9-point scale for each ranging from "very unlikely" to "very likely," with the mid-point labeled "neither likely nor unlikely."

For the frequency estimation task, participants were asked: "Out of 100 Sacramento Ants [Albany Ferns/Rwandan Marmots], how many do you think have each of the following properties?" The four features were then listed in a random order, and participants indicated a number between 0 and 100.

Finally, for the future inference judgments, participants received the following prompt: "For each feature below, please rate how likely you think it is that future Sacramento Ants [Albany Ferns/Rwandan Marmots] will possess that feature." The features and 9-point scale followed, as for the historical inference judgments.

Participants were presented with the three categories in a random order. The categorization task, historical inference judgments, frequency estimates, and future inference judgments immediately followed the feature descriptions for each category, and were presented in one of four orders, corresponding to a Latin square. The order of these judgments was counterbalanced with the two stimulus sets, yielding a total of eight conditions to which participants were randomly assigned.

2.2. Results

Each set of judgments was analyzed with a repeated-measures ANOVA including two within-subjects factors, feature functionality (2: functional, non-functional) and category (3: ant, fern, marmot), with judgment order as a between-subjects factor (4 orders). To facilitate interpretation of categorization judgments in this and subsequent experiments, the 9-point categorization ratings (corresponding to how likely it is that an organism *missing* a given feature is a category member) were subtracted from 10, such that higher ratings reflect greater diagnosticity. Means and standard deviations for each judgment are presented in Table 3 as a function of feature functionality.

Table 2
Sample stimuli from Experiment 1 for the Sacramento Ant category.

Feature	Functional feature description	Non-functional feature description
Sacramento Ants have thick blood	Having thick blood helps the ants cope with parasites, as they bleed less from parasite bites	Having thick blood neither helps nor hurts the ants
Sacramento Ants are red	Being red serves as camouflage to protect the ants from predators	Being red neither helps nor hurts the ants
Sacramento Ants have sticky antennas	Having sticky antennas helps the ants to navigate successfully by ensuring their antenna sensors do not dry out	Having sticky antennas neither helps nor hurts the ants
Sacramento Ants have a slow digestive system	Slowed digestion helps the ants survive when food supplies are short	Slowed digestion neither helps nor hurts the ants

As expected, organisms missing a functional feature received a lower classification rating than those missing a nonfunctional one, reflecting functional features' greater importance in establishing category membership, $F(1,76) = 67.76$, $p < .001$, partial $\eta^2 = .471$. Importantly, however, function also influenced each of the other dependent variables: Functional features were judged more likely to have resulted from evolutionary processes, $F(1,76) = 68.60$, $p < .001$, partial $\eta^2 = .474$, more frequent in the current population, $F(1,76) = 21.79$, $p < .001$, partial $\eta^2 = .223$, and more likely to persist in future populations, $F(1,76) = 81.03$, $p < .001$, partial $\eta^2 = .516$ (see Table 3). The only additional significant effect for any analysis was an interaction between feature functionality and judgment order for the categorization ratings, $F(3,76) = 4.82$, $p < .01$, partial $\eta^2 = .160$. The functional feature was rated significantly more diagnostic of category membership than the nonfunctional feature for all judgment orders, but the difference between these ratings was smallest when the categorization judgments were presented first, with a rating of 7.03 (SD = 1.72) for the functional feature versus 6.35 (SD = 1.72) for the nonfunctional feature.

2.3. Discussion

Experiment 1 established two important findings. First, a category member's functional features are judged more diagnostic of category membership than are non-functional features. For example, an ant's red color is judged more diagnostic when red serves as camouflage. This result confirms the presupposition that motivates this paper, namely, that functional features have a privileged role in biological kind classification. It also extends previous findings demonstrating the importance of functional features for category membership (e.g., Kemler-Nelson et al., 1995; Kim & Rehder, 2011; Lin & Murphy, 1997; Lombrozo, 2009; Murphy & Wisniewski, 1989; Wisniewski, 1995).

The second finding is that functional features license additional inferences: They are judged more likely to have resulted from evolutionary processes, more frequent in the current population, and more likely to persist in the future. That functional features were viewed as resulting from evolutionary processes is consistent with the historical hypothesis. That they were viewed as more prevalent in current and future populations is consistent with the frequency and stability hypotheses. These data are among the first to shed light on the question of why functional features have a special influence on people's classification decisions (see also Ahn, 1998).

As mentioned in the introduction, however, the findings from this initial experiment must be interpreted with caution. The factors that we identify – namely history, frequency, and persistence – could *reflect* a features' status in categorization rather than being a cause of it. For example, a feature could be judged frequent because it is judged important for judging category membership, and not the reverse. Accordingly, the following experiments aim to establish whether these variables are causally implicated in function's effect on classification through experimental manipulation.

3. Experiment 2: Functions and causal history

The aim of Experiment 2 was to assess whether functional features influence categorization because they are assumed to have a deep causal history. Participants were presented with the same three biological kinds with four features as in Experiment 1, but the four features were described as either functional or nonfunctional and as the result of either deep, phylogenetic processes (e.g., natural selection) or more recent and typically ontogenetic processes (e.g., a recent change in the environment). For example, the Sacramento Ant's red color could serve as camouflage as a result of natural selection (functional + phylogenetic), serve as camouflage as a side-effect of recent exposure to high levels of UV (functional + ontogenetic), not have a function but result as a side-effect of a property that is itself a result of natural selection (nonfunctional + phylogenetic), or not have a function and result as a side-effect of recent exposure to high levels of UV (nonfunctional + ontogenetic).

According to the historical hypothesis, functional features are privileged because functionality is a cue to a feature's causal (evolutionary) history that in turn is a cue to the item's category membership. Experiment 1 established the first part of this claim by showing that functional features are indeed viewed as having a deep evolutionary history. If the second part holds as well, then Experiment 2 should yield a main effect of feature history, with phylogenetic features judged more diagnostic of category membership than ontogenetic features. Note that Experiment 2's orthogonal manipulation of history and function allows us to also ask whether causal history fully mediates the function/categorization relationship. If it does, then there should be no main effect of function above and beyond the effect of history. That is, if causal history is controlled, the effect of function should be eliminated.

Finally, if functional features are privileged in biological kinds because of their affordances, as claimed by the affordance hypothesis, then functional features should be judged more diagnostic of category membership (as they were in Experiment 1), and this should be true regardless of whether or not they have a phylogenetic history.

3.1. Methods

3.1.1. Participants

Sixty Berkeley undergraduates or members of the Berkeley community (70% women; mean age 20) participated in exchange for either course credit or pay.[3]

3.1.2. Materials

The stimuli from Experiment 1 were modified as follows. Four features were specified for each category, and each feature could appear in one of four versions: phylogenetic versus ontogenetic crossed with functional versus nonfunctional. There were thus 16 possible feature descriptions for each category (4 features $\times$ 4 versions). The 16 were grouped into four variants for each category of four feature descriptions each, selected such that each category variant involved four unique features and four unique versions. So, for example, if the phylogenetic/functional version of the Sacramento Ant's red color was employed in a given category variant, no other phylogenetic/functional version was employed for that category variant. An example of the stimuli seen by a single participant for the Sacramento Ant category is provided in Table 4; all four versions for a sample feature from the Albany Fern category are provided in Table 5.

3.1.3. Procedure

The study was administered in a lab environment using web-based experiment presentation software. Participants were told that they would learn about three categories of plants or animals. For each category, they read feature descriptions for each of four features. This was followed by categorization judgments and frequency estimation judgments like those in Experiment 1.

Participants saw the three categories in a random order, with the categorization and frequency judgments immediately following the feature descriptions for each category. For each participant, the category variant for a given category was also selected at random, with the constraint that each category variant be selected an equal number of times (15) across participants.

[3] Four additional participants were excluded from analyses for leaving one item or more blank. Including these participants does not affect results, except where noted.

Table 3
Mean judgments for Experiment 1 as a function of feature functionality. Means are followed in parentheses by standard deviations.

Judgment type	Functional features	Nonfunctional features
Categorization rating: Suppose you come across present-day ants [ferns/marmots] that each have three of the four features you just read about, but are missing one of those features. How likely do you think it is that each is a Sacramento Ant [Albany Fern/Rwandan Marmot]? (1–9**)***reverse coded*	7.02 (1.78)	5.45 (1.99)
Frequency estimation: Out of 100 present-day Sacramento Ants [Albany Ferns/ Rwandan Marmots], how many do you think have each of the following properties? (0–100)	86.03 (23.37)	76.98 (26.66)
Historical inference: Some features are primarily the result of evolutionary history, while others are the result of experiences during an organism's recent history (e.g., things that may have changed in the species' environment). For each feature, rate how likely you think it is that the feature results from evolutionary processes. (1–9)	7.50 (1.78)	5.81 (1.90)
Future inference: For each feature, please rate how likely you think it is that future Sacramento Ants will possess that feature (1–9)	8.02 (1.26)	6.20 (1.73)

3.2. Results

Participants' categorization ratings are presented in Fig. 1A as a function of whether features were phylogenetic or ontogenetic and functional or nonfunctional. The ratings were analyzed with a repeated-measures ANOVA including feature history (2: phylogenetic, ontogenetic), feature functionality (2: functional, non-functional), and category (3: ant, fern, marmot) as within-subjects factors. This analysis revealed three main effects. There was a main effect of feature history, $F(1,59) = 25.06$, $p < .01$, partial $\eta^2 = .298$, with phylogenetic features judged more diagnostic than ontogenetic features. There was also a main effect of feature functionality, $F(1,59) = 21.13$, $p < .01$, partial $\eta^2 = .264$, with functional features judged more diagnostic than non-functional features. Finally, there was a main effect of category, $F(2,58) = 3.48$, $p < .05$, partial $\eta^2 = .107$, with lower ratings for the ant category than for the fern or marmot categories. There were no additional effects. Most notably, there was no interaction between feature history or functionality and category, reflecting the fact that all three categories generated the same qualitative pattern of results. Moreover, the effect of feature functionality was significant within both the phylogenetic, $t(59) = 4.74$, $p < .01$, and ontogenetic, $t(59) = 3.05$, $p < .01$, cases.

Frequency ratings were analyzed with an equivalent ANOVA, yielding a main effect of feature history, $F(1,59) = 36.05$, $p < .01$, partial $\eta^2 = .379$, a main effect of feature functionality, $F(1,59) = 12.93$, $p < .01$, partial $\eta^2 = .180$, and an interaction between feature history and category, $F(2,58) = 4.52$, $p < .05$, partial $\eta^2 = .135$ (see Fig. 1B).[4] Features that were phylogenetic were estimated to be more frequent than those that were ontogenetic, and those that were functional were rated more frequent than those that were non-functional. These main effects mirror those for categorization. The interaction reflects the fact that the difference between phylogenetic and ontogenetic features was somewhat greater for the marmot and fern categories than for the ant category, but for all three categories the same qualitative pattern of results emerged.

3.3. Discussion

Experiment 1 found that functional features were viewed as more likely to have a deep evolutionary history. Experiment 2 went beyond this initial finding to ask whether history helps explain the influence of functional features on categorization judgments. The answer is that it does, as phylogenetic features were treated as more diagnostic of category membership than ontogenetic features. For example, where Experiment 1 found that a feature of Albany Ferns like "speckled fronds" was judged more likely to have a deep evolutionary history when it served a function (e.g., attracting but-

[4] The interaction between feature history and category was marginal when the four participants who left items blank were included in the analysis, $F(2,62) = 2.56$, $p = .085$, partial $\eta^2 = .076$.

terflies), Experiment 2 found that speckled fronds were more diagnostic of Albany Ferns when they arose from natural selection versus a recent change in exposure to light.

Besides enhancing our understanding of functional features, the effect of causal history on categorization is a substantive finding in its own right. Of course, other theorists have argued that "deeper" features are more diagnostic of category membership (Ahn et al., 2000; Sloman et al., 1998). However, "depth" in these past studies has referred to a position in a network of dependency or causal relations between *current* features rather than to historical causal processes associated with a category's etiology. Other researchers have also noted how classification can sometimes be an act of inference in which the features that one observes in an item are used to infer the presence of other properties that in turn help establish its category membership (e.g., Gelman, 2003; Medin & Ortony, 1989; Murphy & Medin, 1985; Rehder, 2007; Rehder & Kim, 2010), but again empirical demonstrations of this idea have been limited to the inference of invisible but current "essential" or defining properties of the item (Hampton et al., 2007; Rehder, 2003; Rehder & Kim, 2009, 2010). The results of Experiment 2 suggest that this notion may also apply to the inference of causal processes operating in the past. Our results are in the spirit of each these previous proposals, but they nonetheless represent a substantive and novel contribution.

Although the present data support the causal history hypothesis, recall that our orthogonal manipulation of history and function allowed us to additionally examine whether history fully mediates the relationship between function and classification established in Experiment 1. The answer is that it does not, as features' function continued to influence category membership even controlling for causal history. Thus, the possibility remains that one or more of the additional factors we identified earlier also contributes to function's influence on classification. For example, functional features could have been more diagnostic of category membership even controlling for causal history because they were viewed as more prevalent in current category members, as predicted by the frequency hypothesis. Indeed, support for this possibility comes from the finding that participants in Experiment 2 rated functional features as more frequent in category members. Experiment 3 assesses this potential role of frequency. Another possibility is that ontogenetic functional features were judged more diagnostic because they were deemed more likely to become targets of natural selection and therefore persist in the future, consistent with the stability hypothesis.[5] Experiments 4 and 5 assess the role of future persistence in the effect of feature functionality on classification.

Before moving on, however, we sought to clarify the influence of causal history on classification by asking whether this effect was itself mediated by frequency. This possibility is suggested by the finding that participants judged phylogenetic features not only more diagnostic but also more frequent than ontogenetic features. Accordingly, we replicated Experiment 2 with an important modification: Each feature was specified as holding for 75% of category members. This experiment, which we will refer to as Experiment 2b, revealed a main effect of causal history.[6] That this effect obtained even controlling for frequency indicates that causal history has an independent influence on classification – that is, one that is not fully mediated by frequency. Experiment 3 now goes on to test how frequency itself impacts classification, and whether causal history and frequency can jointly explain the effect of function on classification.

[5] We thank an anonymous reviewer for raising this possibility.

[6] Experiment 2b involved 160 Berkeley undergraduates or members of the Berkeley community (64% women, mean age 20) who each provided categorization ratings for a single category with feature descriptions identical to Experiment 2, except that the frequency of the feature was specified as 75%. Categorization ratings were analyzed with a mixed ANOVA with feature functionality (2: functional, nonfunctional) and feature history (2: phylogenetic, ontogenetic) as within-subjects factors and category (3: ant, fern, marmot) as a between-subjects factor. This analysis revealed, a main effect of feature functionality, $F(1,165) = 4.71$, $p < .05$, partial $\eta^2 = .028$, reflecting the fact that functional features were judged more diagnostic than nonfunctional features, and a main effect of feature history, $F(1,165) = 16.28$, $p < .001$, partial $\eta^2 = .090$, reflecting the fact that phylogenetic features were judged more diagnostic than ontogenetic features. There was also a significant interaction between feature history, feature functionality, and category, $F(2,165) = 4.40$, $p < .05$, partial $\eta^2 = .051$, reflecting the fact that for the fern category only, the item missing the phylogenetic/functional feature received a lower categorization rating than the item missing the ontogenetic/functional feature (5.45 versus 5.50); we attribute this effect to random variation.

Table 4
Sample stimuli from Experiment 2, illustrating one variant for the Sacramento Ant category. Participants assigned to this variant received all four of these feature descriptions.

Feature functionality	Feature history	
	Phylogenetic	Ontogenetic
Functional	Thick blood. Sacramento Ants have thick blood because they evolved in an environment with microscopic parasites. Ants with thicker blood lost less blood from the parasite bites because their blood coagulated more quickly, increasing their chance of survival. As a result of natural selection, most present-day Sacramento Ants have thick blood. Having thick blood continues to help the ants cope with parasites.	Sticky antennas. Sacramento Ants have sticky antennas due to a recent toxic waste leak in the lake that supports the Sacramento Ant population. The toxic waste caused a genetic mutation that affected the secretions on ant antennas. As a result, most present-day Sacramento Ants have sticky antennas. It turns out that having sticky antennas is a useful property, because ants with stickier antennas are less likely to have their antenna sensors dry out and cease to function, and hence can reliably navigate back to the colony.
Non-functional	Red color. Sacramento Ants are red because of the particular species of aphids they eat. Eating the aphids improved the fertility of Sacramento Ants. As a result of natural selection, most present-day Sacramento Ants seek out and eat these aphids, which as a side effect makes the ants red. Being red itself neither helps nor hurts the ants.	Slow digestion. Sacramento Ants have a slow digestive system because of the recent availability of human food sources, which they have learned to eat. Preservatives in the food reduce the production of enzymes that promote digestion. As a result, most present-day Sacramento Ants have a slow digestive system. Having a slow digestive system neither helps nor hurts the ants.

4. Experiment 3: Functions and frequency

The aim of Experiment 3 is to assess whether functional features influence categorization because they are viewed as being more frequent among current category members. Participants were instructed on the same categories as in the first two experiments. Features were described as either functional or nonfunctional and either very common in category members (present in 90% of members) or less common (present in only 60%). To determine whether any effect of frequency holds above and beyond causal history, we controlled for this variable by describing each feature as ontogenetic (e.g., a side-effect of recent changes in the environment). Thus, Experiment 3 also asks whether causal history and frequency together fully explain the effect of function on categorization. If they do, Experiment 3 should yield no main effect of whether features are functional or nonfunctional.

The experiment has two additional but secondary aims. The first is to clarify the role that within-category frequency plays in mediating the relationship between a feature's function and its diagnosticity. Experiments 1 and 2 solicited judgments concerning a feature's frequency within, say, Sacramento Ants, but a feature's diagnosticity is additionally a function of its frequency within *other* categories (e.g., other species of ants). It is possible that being functional leads participants to believe that a feature is not only more frequent among category members (as shown in the first two experiments), but among *non*members as well (Rosch & Mervis, 1975; Sloman et al., 1998). Such an effect would complicate the interpretation of the findings we've presented so far. If functional features are more frequent among both members and nonmembers, frequency no longer explains their greater diagnosticity. Conversely, functional features could be viewed as more diagnostic not because they are *more common* among members, but rather because they are viewed as *less common* among nonmembers. To address these possibilities, Experiment 3 includes an additional post-test question that asks participants to estimate the frequency of each feature among nonmembers.

Finally, Experiment 3 re-introduces the future inference judgments from Experiment 1. Recall that Experiment 1 found that functional features were judged more likely to persist in future populations than nonfunctional features. Experiment 3 therefore considers whether functional features are judged more likely to persist in the future even controlling for causal history and frequency. This measure is relevant for assessing the stability hypothesis, which is the focus of Experiments 4 and 5.

Table 5
Sample stimuli from Experiment 2 for the feature "bitter tasting" from the Albany Fern category. Each participant was presented with a single description from this set.

Feature functionality	Feature history	
	Phylogenetic	Ontogenetic
Functional	Albany Ferns are bitter tasting because they evolved in an environment with fern-eating animals. Bitter-tasting ferns were less likely to be eaten, increasing their chance of survival. As a result of natural selection, most present-day Albany Ferns are bitter tasting. Their bitter taste continues to protect Albany Ferns from being eaten.	Albany Ferns are bitter tasting because recent increases in UV levels have influenced the expression of genes involved in nutrient storage. Some nutrients are consequently stored in a different form, and result in a bitter taste. As a result of this change, most present-day Albany Ferns are bitter tasting. It turns out that tasting bitter is a useful property, because it prevents Albany Ferns from being eaten by animals, increasing their chance of survival.
Non-functional	Albany Ferns are bitter tasting because they have high levels of magnesium. Magnesium plays an important role in photosynthesis, so ferns with more magnesium were more likely to survive. As a result of natural selection, most Albany Ferns have high levels of magnesium. As a side-effect, most present-day Albany Ferns taste bitter. Tasting bitter neither helps nor hurts the ferns.	Albany Ferns are bitter tasting because recent increases in UV levels have influenced the expression of genes involved in nutrient storage. Some nutrients are consequently stored in a different form, and result in a bitter taste. As a result of this change, most present-day Albany Ferns are bitter tasting. Tasting bitter neither helps nor hurts Albany Ferns.

4.1. Methods

4.1.1. Participants

Sixty Berkeley undergraduates (77% women; mean age 20) participated on-line in exchange for course credit.[7]

4.1.2. Materials

Experiment 3 employed the same three novel categories as Experiment 2. However, only the ontogenetic feature descriptions were used, and the descriptions specified that each feature was present in either 60% or 90% of present-day members of the category. In addition, the feature descriptions were modified to equate features more closely: Each was described as having a consequence, and that consequence was either functional or non-functional.

4.1.3. Procedure

The procedure was similar to Experiment 2, but because the frequency of each feature was specified in the feature descriptions, the frequency estimation task was not included.

In addition, after the final category, participants provided two sets of judgments for that category: nonmember frequency estimates and future inference judgments. For the nonmember frequency estimates they received the following prompt: "Out of 100 ants [ferns/marmots] that are NOT Sacramento Ants [Albany Ferns/Rwandan Marmots] – that is, that are other kinds of ants [ferns/marmots] – how many do you think have each of the following features?" Participants entered a number between 0 and 100 for each feature. The future inference prompt was like that in Experiment 1.

Participants were randomly assigned to one of the four category variants for each of the three categories, with the constraint that an equal number of participants (15) complete each category variant.

4.2. Results

Participants' categorization ratings were analyzed with a repeated-measures ANOVA including the following within-subjects factors: feature frequency (2: 60%, 90%), feature functionality (2: functional,

[7] Four additional participants were excluded from analyses for leaving one item or more blank. Including these participants does not affect results, except where noted.

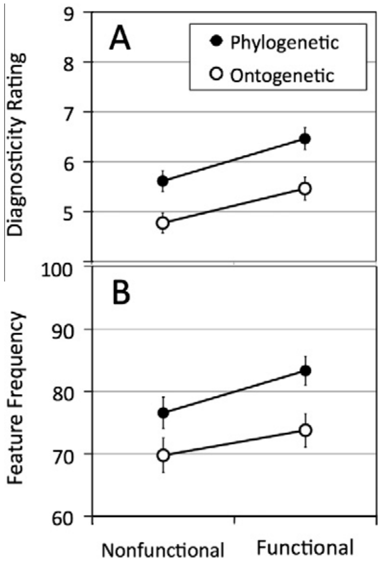

Fig. 1. Average judgments for each feature type in Experiment 2: (A) diagnosticity ratings and (B) feature frequency estimates. Error bars indicate one standard error of the mean in each direction.

non-functional), and category (3: ant, fern, marmot). This analysis revealed a main effect of feature frequency, $F(1,59) = 27.67$, $p < .001$, partial $\eta^2 = .319$, with features present in 90% of category members judged more diagnostic than those present in 60% of members (see Fig. 2A). There was also a main effect of feature functionality, $F(1,59) = 8.59$, $p < .01$, partial $\eta^2 = .127$, with functional features more diagnostic than non-functional features. There were no other significant effects, suggesting that frequency and functionality had comparable effects across conditions. The effect of feature functionality was independently significant within each frequency condition (60%: $t(59) = -2.08$, $p < .05$; 90%: $t(59) = -2.50$, $p < .05$).

The two post-test measures, nonmember frequency and future inference judgments, were only collected for the final category for each participant, making category a between-subjects factor for the purposes of analyzing these judgments. A mixed ANOVA with category as a between-subjects factor (3: ant, fern, marmot), feature frequency (2: 60%, 90%) and feature functionality (2: functional, non-functional) as within-subjects factors, and nonmember frequency judgment as a dependent measure did not reveal any significant effects (see Fig. 2B). The mean judgment for the frequency of a given feature in 100 non-members was 48.06 (SD = 24.19), which did not differ from 50, $t(59) = -.620$, $p = .537$. Assuming that participants treated the scale midpoint as a neutral value, the findings suggest that many participants did not have strong intuitions about the frequency of a given feature among nonmembers.

Future inference judgments were analyzed with an equivalent ANOVA, revealing a main effect of feature functionality, $F(1,57) = 10.66$, $p < .01$, partial $\eta^2 = .158$ (see Fig. 2C). Functional features were judged more likely to persist in the future than were non-functional features. Highly frequent features (90%) were also judged more likely to persist than moderately frequent features (60%), but this trend did not reach significance, $F(1,57) = 3.28$, $p = .075$, partial $\eta^2 = .054$.[8] There were no other significant effects.

[8] The effect of feature frequency on future inference was significant when the four participants excluded for leaving items blank were included in the analysis, $F(1,61) = 4.37$, $p = .04$, partial $\eta^2 = .067$.

4.3. Discussion

Experiments 1 and 2 found that functional features were judged to be more prevalent in current category members. Experiment 3 went beyond these findings to consider whether frequency helps explain those features' influence on categorization judgments. The results reveal that it does, as more frequent features were treated as more diagnostic of category membership than less frequent ones. That this result obtained even controlling for causal history suggests that frequency is a second variable that mediates the influence of a feature's functionality on categorization judgments.

To illustrate the role that causal history and frequency have in mediating function's effect on classification, it is informative to examine effect sizes over the first three experiments. Whereas the difference between the classification ratings of functional and nonfunctional features was 1.57 in Experiment 1 (and the η^2 associated with function was .471), that difference dropped to 0.77 (η^2 = .254) in Experiment 2 when causal history was controlled. In the current experiment, that difference dropped further to .32 (η^2 = .127). These findings must be interpreted with caution of course, as the experiments varied in multiple ways, but bolster our claim that history and frequency mediate the effect of function on categorization.

The present results also solidify the link between features' within-category frequency and their diagnosticity by showing that feature functionality has no effect on features' perceived frequency among *nonmembers* of the same superordinate class. We raised the possibility that differences in feature diagnosticity found in the previous experiments might have been influenced by assumptions about feature frequency among nonmembers. However, participants in Experiment 3 found functional and nonfunctional (and frequent and infrequent) features differentially diagnostic while simultaneously rating those features as equally frequent in other categories.

Experiments 2 and 3 not only support the historical and frequency hypotheses, respectively, they also support the stability hypothesis and indeed provide some initial reasons to favor it over the two other hypotheses. Recall that the stability hypothesis states that functionality licenses inferences about features' frequency among past, present, and future category members and that this inference about the category's (temporally extended) distribution of features influences features' conceptual importance. The present experiment demonstrated the importance of current frequency, and Experiment 2 demonstrated the importance of past frequency (on the assumption that features with a deeper causal history are also viewed as more prevalent in past category members). Further, Experiment 3 additionally found that features with functions were judged more likely to persist in the future, and this effect obtained even controlling for causal history and current frequency. This result suggests that the residual effect of functional features found in the current experiment might be due to inferences concerning the persistence of functional features into the future. Experiment 4 directly investigates this possibility by assessing the impact of feature functionality and a feature's persistence on judgments of category membership.

5. Experiment 4: Functions and the future

The aim of Experiment 4 was to assess whether functional features influence categorization because they are stable over time. As Experiment 2 manipulated a feature's past and Experiment 3 its frequency in the present, Experiment 4 turns to the future by varying a feature's persistence in future populations. Specifically, features were described as either temporary or permanent. For example, the Sacramento Ant's red color always resulted from an ontogenetic process: a recent exposure to high levels of UV light. However, the exposure was either described as temporary, in which case future Sacramento Ants would not be red, or permanent, in which case future Sacramento Ants would continue to be red. A feature's permanence was varied independently of its functionality, such that the red color was described as functional (camouflage) for some participants and nonfunctional for others. As in Experiment 3, we controlled causal history by describing each feature as ontogenetic. According to the stability hypothesis, permanent features should be more diagnostic of category membership than temporary ones.

5.1. Methods

5.1.1. Participants

Eighty-four Berkeley undergraduates or members of the Berkeley community (77% women; mean age 20) participated in exchange for either course credit or pay.[9]

5.1.2. Materials

Experiment 4 employed the same novel categories as the previous experiments, with feature descriptions similar to those in Experiment 3. However, the frequency of each feature was not specified, and each feature had a consequence that was either temporary or permanent. See Table 6 for sample features.

5.1.3. Procedure

The procedure mirrored the previous experiments in that participants were presented with all three novel categories and completed a categorization task and a frequency estimation task for each category. In addition, participants completed a post-test for the final category requesting judgments about each feature's functionality and persistence in the future. The post-test serves as a manipulation check to ensure that participants understood and accepted the claims in the feature descriptions about permanence. For the functionality judgments, participants were asked to rate how important they thought each feature was for the category members' survival on a 9-point scale ranging from "not at all important" to "very important," with "neither important nor unimportant" as a midpoint. The future inference judgments were like those in Experiments 1 and 3, with participants asked how likely they thought it was that future members of the category would possess the feature in question, with a 9-point scale ranging from "very unlikely" to "very likely," with "neither likely nor unlikely" as a midpoint.

Participants were randomly assigned to one of three category orders conforming to a Latin square, with 28 participants per order. Category variants were selected at random, with the constraint that the same number of participants (21) completed each variant.

5.2. Results

Categorization judgments were analyzed with a repeated-measures ANOVA including feature permanence (2: permanent, temporary), feature functionality (2: functional, non-functional), and category (3: ant, fern, marmot) as within-subjects factors. This analysis revealed a main effect of permanence, $F(1,83) = 29.63$, $p < .001$, partial $\eta^2 = .263$, with permanent features judged more diagnostic than temporary features, and a main effect of feature functionality, $F(1,83) = 11.48$, $p < .01$, partial $\eta^2 = .121$, with functional features judged more diagnostic than nonfunctional features (see Fig. 3A). The effect of functionality on classification was significant for features that were temporary, $t(83) = -2.88$, $p < .01$, as well as those that were permanent, $t(83) = -2.94$, $p < .01$.

An equivalent ANOVA with frequency estimates mirrored these results, revealing a main effect of feature permanence, $F(1,83) = 26.85$, $p < .01$, partial $\eta^2 = .244$, as well as a main effect of feature functionality, $F(1,183) = 25.01$, $p < .01$, partial $\eta^2 = .232$ (see Fig. 3B). No other effects concerning categorization or frequency were statistically significant.

The two post-test measures, functionality and future inference judgments, were only collected for the final category for each participant, making category a between-subjects factor for the purposes of analyzing these judgments. A mixed ANOVA with category as a between-subjects factor (3: ant, fern, marmot), feature permanence (2: permanent, temporary) and feature functionality (2: functional, nonfunctional) as within-subjects factors, and functionality judgment as a dependent measure revealed a main effect of feature functionality, $F(1,81) = 14.09$, $p < .01$, partial $\eta^2 = .148$, as well as an interaction between feature functionality and category, $F(2,81) = 6.89$, $p < .01$, partial $\eta^2 = .145$. The

[9] An additional 15 participants were replaced for leaving one item or more blank. However, the results are not affected by including these participants in analyses.

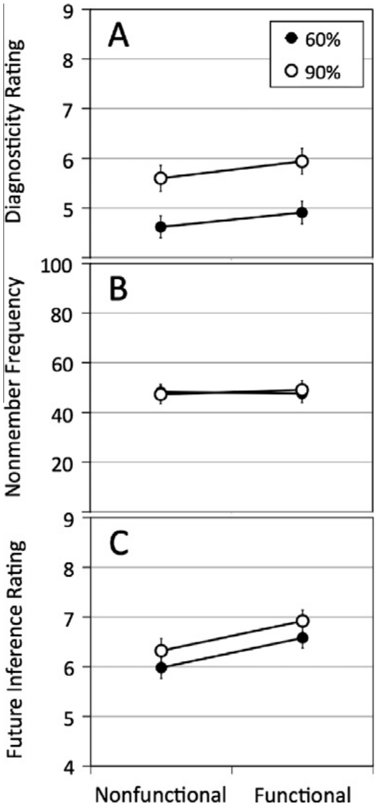

Fig. 2. Average judgments for each feature type in Experiment 3: (A) diagnosticity ratings, (B) nonmember frequency estimates, and (c) future inference ratings. Error bars indicate one standard error of the mean in each direction.

main effect of feature functionality serves as a manipulation check, and reflects the fact that functional features were judged more important for survival than nonfunctional features (see Fig. 3C). The interaction reflects the fact that functionality ratings differed substantially between functional and nonfunctional features in the ant and fern categories, but not in the marmot category.

Future inference judgments were analyzed with an equivalent ANOVA, revealing a main effect of feature permanence, $F(1,81) = 7.11$, $p < .01$, partial $\eta^2 = .081$, as well as a main effect of feature functionality, $F(1,81) = 6.70$, $p < .01$, partial $\eta^2 = .076$ (see Fig. 3D). The effect of feature permanence serves as a manipulation check, and indicates that features stipulated to be permanent were judged more likely to persist in the future than those stipulated to be temporary. Interestingly, however, feature functionality had significant effects on future inference ratings as well, with features that served functions – even functions stipulated to be temporary – judged more likely to persist in the future than those without functions.

5.3. Discussion

Whereas Experiment 1 found that functional feature's were viewed as more likely to persist in the future, Experiment 4 asked whether that persistence helps explain those features' greater influence on categorization judgments. The answer is that it does, as permanent ontogenetic features were treated as more diagnostic of category membership than temporary ones. This relationship held even when features' causal history was controlled.

Another finding in Experiment 4 was that there was still an effect of function on classification even controlling for causal history and permanence. On one hand, this result could be interpreted as supporting the affordance hypothesis. On the other hand, Experiment 4 also found that functional features were judged to be more prevalent in the current population of category members and we saw (in Experiment 3) that feature frequency influences classification. To determine whether the residual effect of function in Experiment 4 was due to frequency, we conducted a replication in which frequency was controlled.

6. Experiment 5: Remaining effects of function

Experiment 5 replicates Experiment 4 with one important modification: Each feature is specified as holding for 75% of the population. If the residual effect of function in Experiment 4 was due to inferences about a feature's current frequency, then the effect of function should fail to replicate in Experiment 5. In contrast, the affordance hypothesis predicts that a feature's functionality should continue to influence category membership judgments even when causal history, permanence, and current frequency are all specified.

6.1. Methods

6.1.1. Participants

One-hundred-sixty-eight participants were recruited on-line from Amazon Mechanical Turk (57% women, mean age 32) and paid in exchange for their participation.[10] Using Mechanical Turk's qualification settings, participation was restricted to people in the United States and who had an approval rate of 95% or higher.

6.1.2. Materials

Experiment 5 employed the same materials as Experiment 4, except that each feature description specified that the feature was present in 75% of present-day category members.

6.1.3. Procedure

The procedure was identical to that in Experiment 4 except that the frequency estimation task was removed.

6.2. Results

Categorization judgments were analyzed with a repeated-measures ANOVA including feature permanence (2: permanent, temporary), feature functionality (2: functional, non-functional), and category (3: ant, fern, marmot) as within-subjects factors. Like Experiment 4, this analysis revealed a main effect of permanence, $F(1,167) = 23.76$, $p < .001$, partial $\eta^2 = .125$, with permanent features judged more diagnostic than temporary features, and a main effect of feature functionality, $F(1,167) = 10.18$, $p < .01$, partial $\eta^2 = .057$, with functional features judged more diagnostic than non-functional features (see Fig. 4A). There were no other significant effects, and the effect of feature functionality on classification was independently significant for features that were permanent, $t(167) = -2.71$, $p < .01$, and for those that were temporary, $t(167) = -2.24$, $p < .05$.[11]

The two post-test measures, functionality and future inference judgments, were only collected for the final category for each participant, making category a between-subjects factor for the purposes of analyzing these judgments. A mixed ANOVA with category as a between-subjects factor (3: ant, fern, marmot), feature permanence (2: permanent, temporary) and feature functionality (2: functional, nonfunctional) as within-subjects factors, and functionality judgment as a dependent measure

[10] An additional 55 participants were excluded for failing the instructional manipulation check that was also used in Experiment 1 (see footnote 2), and 12 for leaving one item or more blank. However, only one result is affected by including these additional 67 participants: a simple effects test that is flagged with a footnote.

[11] When including the 67 participants who were excluded for failing the instructional manipulation or leaving items blank, the effect of feature functionality within the temporary condition is no longer significant, $t(234) = -1.19$, $p = .24$.

revealed a main effect of feature functionality, $F(1,165) = 39.63$, $p < .001$, partial $\eta^2 = .194$. Functional features were judged more important for survival than non-functional features (see Fig. 4B). There was also a significant effect of category, $F(2,165) = 5.29$, $p < .01$, partial $\eta^2 = .060$, with functionality ratings slightly but consistently lower for ferns than for ants or marmots. There were no other significant effects.

Finally, an equivalent ANOVA on future inference judgments revealed a main effect of feature permanence, $F(1,165) = 27.44$, $p < .001$, partial $\eta^2 = .143$, as well as a main effect of feature functionality, $F(1,165) = 4.67$, $p < .05$, partial $\eta^2 = .028$ (see Fig. 4C), mirroring the findings from Experiment 4. Features that were stipulated as permanent rather than temporary, and those that were functional rather than nonfunctional, were judged more likely to persist in the future. No other effects were significant.

6.3. Discussion

Consistent with Experiment 4, Experiment 5 found an effect of feature permanence such that permanent features had greater influence on classification judgments than temporary ones. That this result obtained controlling for both causal history and current frequency confirms that permanence is yet a third variable that mediates function's influence on categorization judgments.[12]

Experiment 5 also asked whether there would be an effect of function even controlling for causal history, current frequency, and persistence in future populations. The answer is that there was, but just barely. Although this provides *prima facie* evidence in favor of the affordance hypothesis, it is important to note that the effect of function in Experiment 5 was smaller (a difference in classification ratings of only 0.20 on a 1–9 scale, $\eta^2 = .057$) than in any previous experiment: 1.57 ($\eta^2 = .471$) in Experiment 1, 0.77 ($\eta^2 = .254$) in Experiment 2, .32 ($\eta^2 = .127$) in Experiment 3, and .43 ($\eta^2 = .121$) in Experiment 4. It is also possible that this small effect of function arose because our permanence manipulation was not completely effective: In both Experiments 4 and 5, post-test measures revealed that functional features were viewed as more likely to persist than nonfunctional ones even though feature permanence was stipulated. Thus, although a small portion of the effect could be explained by the affordance hypothesis, our experiments show that history, frequency, and permanence account almost entirely for the effect of a feature's functionality on its diagnosticity.

7. Summary of inter-variable relationships

Fig. 5 summarizes the relationships between variables that we believe are licensed by our five experiments. First, Experiment 1 manipulated function and found that it influences causal history, current frequency, and permanence. Next, Experiment 2 manipulated causal history and found that it influences current frequency, but Experiment 2b (see footnote 6) found that causal history affects categorization even when frequency is controlled. Experiment 3 manipulated frequency and found that it influences categorization. Experiment 4 manipulated permanence and found that it affected frequency; Experiment 5 then found that permanence had an effect on categorization even controlling for frequency. Finally, the small, residual effect of function in Experiment 5 leaves open the possibility (depicted as a dotted line in Fig. 5) that function has a direct effect on categorization (the affordance hypothesis) or that this relationship is mediated by some other, as yet unidentified variable.

Although these experiments establish that causal history, frequency, and permanence mediate the relationship between function and categorization, they leave unanswered some questions regarding the relationships among the mediators themselves. On one hand, Experiment 2b established that causal history influences categorization controlling for frequency, Experiment 3 that frequency influences categorization controlling for causal history, and Experiment 5 that permanence influences

[12] It is worth highlighting that both Experiments 4 and 5 were restricted to ontogenetic features. We would anticipate similar effects for phylogenetic features, however, provided that one could avoid ceiling effects for feature permanence. We thank an anonymous reviewer for noting this point.

Table 6
Sample stimuli from Experiment 4 for the feature "oily fur" from the Rwandan Marmot category.

Feature functionality	Feature permanence	
	Permanent	Temporary
Functional	Rwandan Marmots have oily fur because of recent seismic activity. Thick oil and tar have seeped through new cracks in the earth's surface in the areas where the marmots burrow and feed. As a result, most present-day Rwandan Marmots have oily fur and, because the seismic activity is expected to continue indefinitely, most future Rwandan Marmots will too. It turns out that because it repels water, oily fur increases the marmots' body temperature, and the retained body heat increases their chances for survival.	Rwandan Marmots have oily fur because of recent seismic activity. Thick oil and tar have seeped through new cracks in the earth's surface in the areas where the marmots burrow and feed. As a result, most present-day Rwandan Marmots have oily fur, but because the seismic activity is expected to be temporary, this will not be true of future Rwandan Marmots. It turns out that because it repels water, oily fur increases the marmots' body temperature, and the retained body heat increases their chances for survival.
Non-functional	Rwandan Marmots have oily fur because of recent seismic activity. Thick oil and tar have seeped through new cracks in the earth's surface in the areas where the marmots burrow and feed. As a result, most present-day Rwandan Marmots have oily fur and, because the seismic activity is expected to continue indefinitely, most future Rwandan Marmots will too. It turns out that because it repels water, oily fur increases the marmots' body temperature. The raised body temperature neither helps nor hurts the marmots.	Rwandan Marmots have oily fur because of recent seismic activity. Thick oil and tar have seeped through new cracks in the earth's surface in the areas where the marmots burrow and feed. As a result, most present-day Rwandan Marmots have oily fur, but because the seismic activity is expected to be temporary, this will not be true of future Rwandan Marmots. It turns out that because it repels water, oily fur increases the marmots' body temperature. The raised body temperature neither helps nor hurts the marmots.

categorization controlling for causal history and frequency. On the other hand, we never experimentally tested whether causal history and frequency influence categorization controlling for permanence. Fortunately, this possibility can be partially assessed with multiple regression analyses in which effects are controlled statistically rather than experimentally. Regression analyses of Experiment 1 confirmed that history ratings predicted categorization ratings even controlling for permanence, $t(79) = 2.54$, $p < .05$, and similar analyses of Experiments 1, 3, and 4 found that frequency predicted categorization ratings controlling for permanence, all ps < .01.[13] These results support the claim that causal history, frequency, and permanence each independently mediate the influence of function on categorization. Finally, these analyses can also be used to augment the results of Experiment 5 by asking whether function remains a significant predictor controlling for the three mediators. The answer is that it does in Experiments 1 and 4 but not in Experiment 3 (p = .0008, .04, and .48, respectively). These results confirm that there might be a residual effect of function, albeit a small one.

8. General discussion

Five experiments investigated whether and why functional features play a special role in biological kind concepts, where we operationalized a feature's importance as its role in establishing category membership. Experiment 1 found that a biological kind's functional features were judged more diagnostic of category membership. They were also judged more likely to have a deep evolutionary history, to be frequent in the current population, and to persist in the future. Experiments 2–5 examined

[13] The reported analyses consisted of carrying out multiple regressions of the categorization ratings from each participant and then testing whether a predictor's regression weight averaged across participants differed significantly from zero. Analysis of Experiment 1 included ratings from all three categories presented to participants whereas those of Experiments 3 and 4 only included those from the final category (the only category for which participants generated permanence ratings and frequency ratings, respectively). Frequency predicted categorization ratings controlling for permanence in Experiment 1, $t(79) = 4.01$, $p < .001$, Experiment 3, $t(59) = 4.89$, $p < .0001$, and Experiment 4, $t(83) = 3.30$, $p < .01$. Frequency and permanence served as categorical predictors in Experiments 3 and 4, respectively. Note that an analysis of the history/permanence relationship could not be performed for Experiments 3 and 4 because those experiments did not gather historical inference judgments.

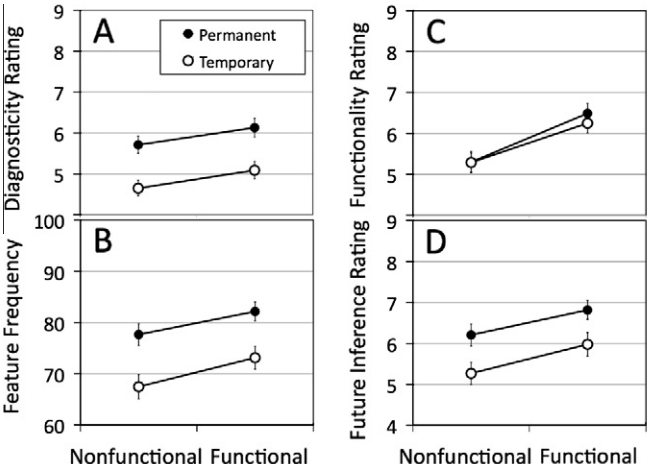

Fig. 3. Average judgments for each feature type in Experiment 4: (A) diagnosticity ratings, (B) feature frequency estimates, (C) functionality ratings, and (D) future inference ratings. Error bars indicate one standard error of the mean in each direction.

whether these additional inferences about functional features mediated the relationship between function and classification. The experiments revealed that they do: Causal history (Experiment 2), frequency (Experiment 3), and permanence (Experiments 4 and 5) accounted almost entirely for the relationship between functionality and classification. Although the experiments failed to completely eliminate effects of function on categorization, the magnitude of the effect of function was substantially decreased with each manipulation, leaving only a small residual effect. This suggests that history, frequency, and permanence jointly account for virtually all of the effect of feature functionality on categorization.

Taken together, the stability hypothesis provides the most parsimonious account of our results. According to this hypothesis, functional features are important to categorization because they are perceived to be frequent not only in the present, but also in the past and in the future, and therefore share a stable relationship to category membership. The effects of causal history and feature frequency support the historical and frequency hypotheses, respectively (see Table 1 and Fig. 5), but the effects of function on inferences about causal history (the past), current frequency (the present), and permanence (the future) are all subsumed by the idea that functional features' association with a category is stable over time and that this association drives their importance to classification.

Why might features that are viewed as stable over time carry special weight in determining category membership? A feature that is more frequent among category members is (all else being equal) normatively more diagnostic of membership, and the link between a feature's frequency and its diagnosticity is of course one of the most well-established findings in categorization research (e.g., Hampton, 1979; Nosofsky, 1988; Rosch & Mervis, 1975). Our proposal goes beyond this previous work, however, in emphasizing that a reasoner's beliefs about the prevalence of a feature in past and future category members is also an important factor determining its diagnosticity. We suggest that stability is important because the purpose of categories is not only to classify, but also to support reasoning more generally, including predictions about the future and explanations that appeal to the past. As these considerations apply quite broadly, an important role for stability is likely to extend beyond biological kinds and functional features. We now consider reasons why functional features, in particular, exhibit such stability.

8.1. From stability to homeostasis

Why might people assume that functional features share a temporally stable relationship to category membership? As mentioned in the introduction, we suspect that this assumption is grounded in the sorts of beliefs that people have about the causal processes associated with functional features. For example, it could be that functional features are conceptualized as participating in causal cycles involving other features and the environment, leading to their perpetuation among category members. To illustrate concretely, consider the feature "red color" for Sacramento Ants, where redness plays a role in camouflage. Suppose further that redness is the result of a constellation of causes C (including genes, pigments, and other factors), and that it generates a series of effects E (including a lower probability of being eaten by predators). So for a given ant, there is an acyclic causal structure in which C causes redness which causes E. Over time, however, causal cycles involving the ants and their environment will play out. At the level of the individual ant, E may play a role in the continued presence of C and thus of redness (for example, because the ant effectively avoids predators and therefore continues to produce the pigments that result in redness). At the level of the category, E will play a causal role in bringing about causes C in the ant's offspring (for example, because it survives long enough to reproduce), and therefore in maintaining and perpetuating the functional feature among category members. Such cycles will characterize functional features in general, and those with historical functions – which result from a consequence-driven process such as natural selection – in particular.

The proposal that functional features are important because they participate in self-perpetuating cycles recalls ideas from philosophy and cognitive development concerning natural kinds. Specifically, Keil (1989, 1995) suggests that natural kind concepts could correspond to clusters of properties that exhibit "causal homeostasis" – reliable co-occurrence brought about by some underlying causal mechanism or regularity. His proposal draws upon philosopher Richard Boyd's notion of a "homeostatic property cluster," an alternative to strong, metaphysical essentialism that nonetheless grounds kinds in real causal properties of the world (Boyd, 1999). For example, Boyd suggests that biological species pick out individuals that share a stable constellation of properties resulting from "homeostatic causal mechanisms." Although the nature of evolutionary change makes it unlikely that there will be a set of necessary and sufficient properties underlying species membership, the stable similarities among species members will nonetheless support useful predictions and explanations. Our proposal can be seen as an extension of these ideas to the role of functional features in biological kind concepts: Functional features could play an especially prominent role in homeostatic clusters, and therefore in classifying biological kinds. This sheds light on why functions are central to the individuation of biological kinds and to the way in which species concepts are structured.

Importantly, our suggestion that people view features with a biological function as participating in causal cycles or homeostatic clusters does not entail that they possess a veridical understanding of evolution. In fact, our results suggest that participants misunderstand natural selection. Recall that participants in Experiments 4 and 5 judged that functional features were more likely to persist than nonfunctional ones even when the causes of those features were stipulated to be temporary. For example, even when participants were told that Rwandan Marmots had oily fur due to temporary seismic activity, they believed that oily fur was (somehow) more likely to persist in the future when it also promoted the marmots' survival by keeping them warm. Our participants seemed to believe that the organism or its environment would (somehow) change to support the perpetuation of (incidentally) functional features. But while our participants' concern for biological kinds may be laudable, their optimism concerning the marmots' future is likely misplaced. Such beliefs are related to prevalent misconceptions about natural selection, and in particular the belief that selective pressures increase the probability of favorable mutations, as opposed to the (more accurate) alternative that selective pressures increase the probability that favorable mutations that already exist will propagate (see Lombrozo, Shtulman, & Weisberg, 2006; Shtulman, 2006). To the extent that people misunderstand evolution as a goal-directed process that operates over individuals, they could be even more inclined to represent functional features as hubs for causal cycles, and thus to privilege them in conceptual representations of biological kinds.

Future research will be required to confirm that people believe functional features participate in the kinds of causal cycles we have described and that those cycles lead classifiers to believe that such features are especially prevalent and stable among category members. Nevertheless, preliminary evidence suggests that causal cycles can indeed increase perceived frequency, at least among current category members. For example, Rehder and Martin (2011) taught participants artificial categories with features that were involved in either causal cycles or acyclic causal structures and found that the for-

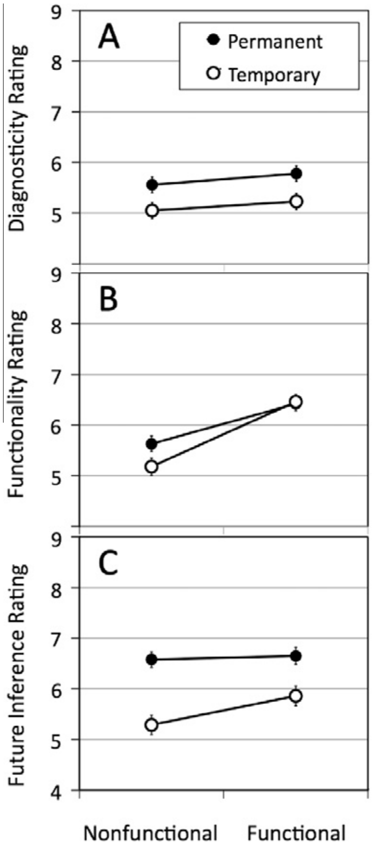

Fig. 4. Average judgments for each feature type in Experiment 5: (A) feature diagnosticity ratings, (B) functionality ratings, and (C) future inference ratings. Error bars indicate one standard error of the mean in each direction.

mer were rated as more frequent in category members than the latter, as well as more diagnostic of category membership. These findings provide preliminary but promising support for our proposal.

8.2. Alternative hypotheses

Our experiments are consistent with the proposal that the frequency of functional features over time makes them especially suitable for establishing category membership, but there are nonetheless a variety of alternative hypotheses worth acknowledging. First, functional features could be privileged because they license functional explanations (e.g., "Sacramento Ants are red for camouflage") and support a functional or "teleological" mode of construal (Keil, 1992, 1994; Kelemen, 1999; Lombrozo & Carey, 2006). Support for this idea comes from Lombrozo (2009), who found that functional features were indeed weighed more heavily in judgments of category membership when they were explained functionally than when they were explained only by appeal to proximate causes. However, previous research also suggests that functional explanations require historical functions (Lombrozo & Carey, 2006). This hypothesis therefore predicts that Experiment 2, which orthogonally varied causal history and functionality, should have generated an interaction, with the feature with a historical function (phylogenetic + functional) judged especially diagnostic because it not only had a phylogenetic history and particular affordances, but supported a functional explanation as well. As this was not observed, the current experiments provide some initial evidence against this "functional explanation" hypothesis.

Another possibility is that functions play a role in establishing an ideal for a category (Barsalou, 1985), with the consequence that category members with functional features are judged better members. To illustrate, consider again the Sacramento Ant, which is red for camouflage. A Sacramento Ant could be closer to the ideal for its category the better its camouflage, and thus if it is red. Although the current experiments were not designed to bear on this hypothesis, the fact that causal history, current frequency, and permanence accounted for the overwhelming majority of the effect of function suggests that if ideals do play a role, it is not a large one. It is also possible that ideals play a larger role for markers of conceptual structure other than feature diagnosticity, such as judgments of category members' typicality (Barsalou, 1985; but see Kim & Murphy, 2011).

Third, it could be that functional features are important as a consequence of their downstream effects, as suggested by Ahn (1998) and discussed in the introduction. In particular. this hypothesis provides a potential alternative explanation for the remaining effect of function in Experiment 5, which we attributed to affordances or to additional, unidentified factors. In all of our experiments, the stimulus materials were matched by having a given feature appear in both a functional and a nonfunctional form across participants. Although this equated the features along many dimensions, the functional versions of the features necessarily had at least one more downstream effect than the nonfunctional features – namely the consequences of the function itself (e.g., avoiding predators for red Sacramento Ants). Given that the factors we manipulated in Experiments 1–5 accounted for the vast majority of the influence of function on classification, however, it seems unlikely that this property of our stimulus materials accounted for much variance in judgments of diagnosticity.

Finally, although we endorse a view in which functional features are central because of their subjective stability, where we have equated stability with elevated frequency over time, we acknowledge that there may be cases in which frequency and stability dissociate. For example, reproductive properties of an organism, such as laying eggs for ducks, are likely to be quite important for category membership even when the relevant features are not overwhelmingly frequent among members (less than 50% of ducks lay eggs; see Carlson & Pelletier, 1995; Leslie, 2011 for additional examples from research on generics). On the flipside, some frequent properties may not be especially important. For example, the vast majority of sheep have their tails shortened ("docked") soon after they are born, a practice that facilitates maintaining and shearing sheep.[14] Nonetheless, having a short tail may not be central to the concept of sheep. Cases like these can be understood by assuming that functional features are central because the causal cycles in which they participate render them not just frequent over time, but also *robust* (or *immutable*; Sloman et al., 1998; see also Lombrozo, 2010) in the face of changing conditions. Future research could aim to differentiate robustness from frequency by, for example, assessing how functional features respond to counterfactual changes to their environment even when frequency over time is controlled.[15] For example, an ant whose redness provides camouflage could be judged more likely than an incidentally red ant to undergo a color change when moved to a green environment (where red is poor camouflage), but *less* likely to undergo a color change when moved to a hotter environment (where camouflage is unaffected). In other words, it could be that a feature's stability is a typical *symptom* of conceptual importance and not *constitutive* of conceptual importance.

8.3. From biological kinds to artifacts

Our experiments are among the first to systematically examine the role of feature functionality on classification within the domain of biological kinds. As discussed in the introduction, there has been considerable research on the role of functions for artifact categories (e.g., Laurence & Margolis, 2007), but there are a few important ways in which the role of functions could differ across these domains. First, for artifacts, both entire objects and object features can have functions: A car is for transportation, and its mirrors are to improve visibility for the driver. In contrast, functions for biological kinds

[14] The first author thanks her Australian husband for the example.

[15] Sloman et al. (1998) also proposed that mutability was dissociable from frequency (what they referred to as *variability*). Nevertheless, empirical demonstrations of such dissociations are not easy to come by. For example, Sloman et al.'s Study 5 found that decreasing a feature's mutability (by manipulating its causal relations) increased its importance for categorization, but also its subjective frequency, supporting our suggestion that the former is often mediated by the latter.

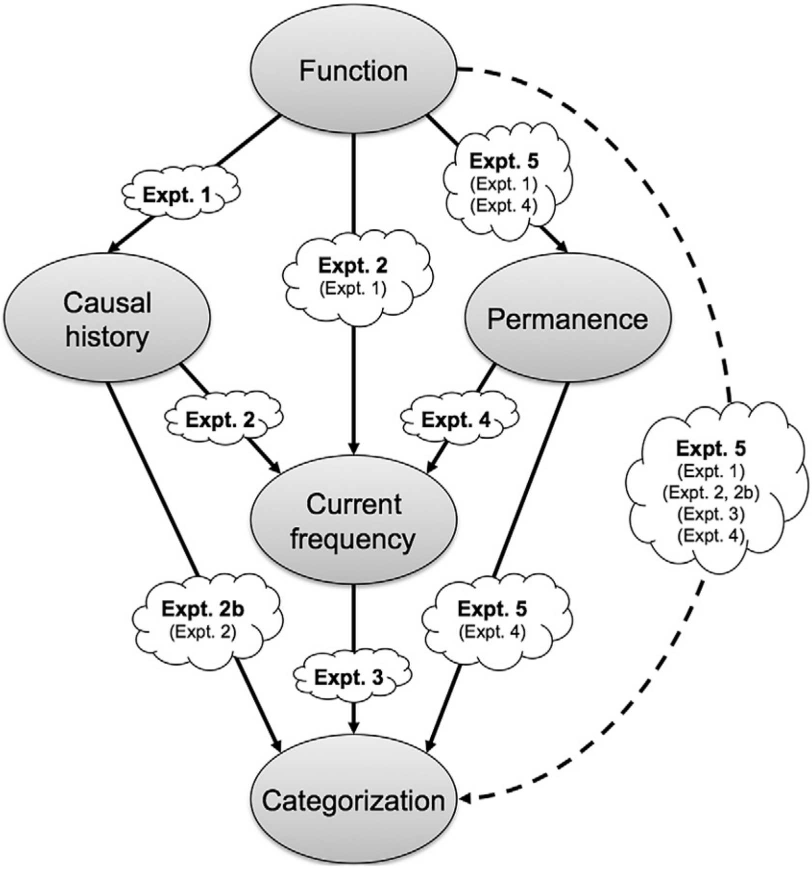

Fig. 5. A graphical depiction of the relationships between the variables examined in Experiments 1–5, including causal relationships between variables. The bubbles along causal arrows identify the experiment that establishes the corresponding relationship. Where more than one experiment is relevant for a given link, we indicate first and in bold the experiment that provides the strongest evidence for a direct relationship (i.e., one that is not mediated by other variables that we consider).

only make sense at the level of features: A zebra's coloration is for camouflage, but the zebra itself is not "for" anything (see, however, Kelemen (1999) for children's judgments, and Greif, Kemler Nelson, Keil, and Gutierrez (2006) for an alternative perspective). It could be that functions at the level of entire objects or even entire superordinate categories (e.g., vehicles) play an important role in artifact concepts without an analog for biological kinds. On the other hand, domesticated species and some concepts that apply to groups of biological kinds, such as predators, weeds, or fruit-bearing trees, could be an interesting hybrid between biological and artifact concepts with respect to the role of functions (see Lynch, Coley, & Medin, 2000, for suggestive though indirect evidence to this effect).

Second, although features from categories within both domains can involve historical functions, those for artifacts are likely to involve a designer's intentions rather than natural selection (for discussion see Lombrozo & Carey, 2006). Moreover, it is unclear whether such intentions are treated as analogous to natural selection when it comes to classification. For example, Chaigneau et al.'s (2004) HIPE theory (also see Barsalou et al., 2005) assumes a mental representation of an artifact in which its designer's intention is viewed as responsible for its physical structure, which in turn is responsible (in conjunction with the actions of an agent) for a functional outcome. According to this approach, the influence of intentions on classification is only to allow the classifier to infer physical properties when they are otherwise unobserved. So, for example, if it is unknown whether an object has a seat, then knowledge that it was designed to be a chair can increase one's belief that it is one because the presence of a seat is inferred. However, more recent studies find that intentions may be an independent source of evidence of category membership (Chaigneau, Castillo, & Martínez, 2008; Puebla-Ramirez, 2011), suggesting that historical functions could be important to artifacts just as they are to biological kinds.

Finally, a straightforward analogy between intentions and natural selection may also fail to hold if the relevant intention for artifact classification does not correspond to an intention to achieve a particular function. Indeed, some accounts have emphasized a designer's intention to create something that belongs to a given category (Bloom, 1996, 1998). For example, the (successfully executed) intention to create a "chair" could be sufficient for an object to be categorized a chair, even without the intention that it successfully support sitting (perhaps it's a chair made of cotton candy). Natural processes such as evolution do not provide an analog for this kind of intention. Unless people conceptualize biological kinds as the products of intentional design (which could be the case for some creationists), this feature of artifact concepts presents another disanalogy between artifacts and biological kinds.

Despite these differences across domains, we expect that our central insights concerning the conceptual role of functional features are likely to generalize to artifacts. In particular, functional features in artifact categories could be privileged because they are believed to share a temporally stable (and possibly a robust and immutable) relationship to category membership. As in biological kinds, this belief could be the result of the causal role that functional features play, including self-reinforcing causal processes that involve the artifact's design process and physical realization.

8.4. Additional limitations and extensions

Our stimulus materials involved a range of biological functions (e.g., coloring to hide from predators or attract pollinators, minerals for blood coagulation or photosynthesis), but we did not systematically manipulate the nature of the functions nor their importance for the organism. One could, for example, vary whether a function concerns survival versus reproduction, the prevention of something harmful versus the generation of something beneficial, or factors internal versus external to the organism. Functions could also vary in their efficacy and necessity, and represent intermediate possibilities between clear historical and ahistorical functions, such as what are sometimes called "pre-adaptations." Taking a more fine-grained look at biological functions and their role in classification could potentially inform specific hypotheses about the causal beliefs that underlie a commitment that functional features share a stable relationship to category membership, and additionally reveal whether serving a function is treated as a qualitatively distinct property of a feature and/or valued as a matter of degree, with factors like those identified above influencing the magnitude of a functional feature's role in classification.

Another direction for future research concerns the way in which a feature's conceptual importance is assessed. Our tasks examined relatively explicit judgments concerning category membership as well as inferences from features with a function that was clearly identified or stipulated not to exist. It's possible that other markers of conceptual structure, such as typicality, could yield different patterns of judgments, or that functions play a less prominent role when specified more subtly. Our own suspicion is that effects of function will be relatively consistent and emerge on more implicit measures in addition to those employed here, but this remains an open empirical question.

Finally, there are interesting questions concerning the extent of individual and cultural variation in the role of functions in biological classification. We anticipate that the factors that contribute to biological classification identified here will hold quite broadly, but past research provides reasons to anticipate at least some variation. In particular, preferences for functional explanations are known to vary both developmentally (e.g., Kelemen, 1999) and cross-culturally (e.g., Casler & Kelemen, 2008), with some individuals accepting functional explanations that others would reject (e.g., "lions are for going in the zoo"). It is plausible that those with a greater preference for functional explanations will tend to weight functional features more heavily in judgments of category membership (Lombrozo, 2009). There is also evidence that folk taxonomic classification varies cross culturally (e.g., Atran, 1994), as does categorization, more generally, as a function of context and the goals of the reasoner (e.g., Ross, 1997). Variation in individuals' goals and expertise could thus translate into subtly different roles for functional information. Even within the biological sciences, there's controversy concerning the appropriate role for functions across taxonomic systems, and this variation could track differences in goals with analogs outside academic science. Future work could capitalize on this variation by examining the role of functional features in judgments across different populations of both expert and novice reasoners.

9. Conclusions

We have shown that if a feature of a biological kind has a function, then that feature is judged more diagnostic of category membership, more likely to have a deep evolutionary history, more likely to be frequent among current category members, and more likely to persist in the future. We have also shown that these inferences about a feature's past, present, and future account almost entirely for effects of feature functionality on classification. We propose that functional features are privileged in representations of biological kinds because they are causally interwoven with other features of the organism and the environment and with the category's history in a way that suggests a stable relationship to category membership over time. This proposal has implications for folk biological understanding, theories of conceptual representation, and models of classification.

Acknowledgments

We thank the Concepts and Cognition Lab at UC Berkeley and the Concepts and Categories Lab at NYU for valuable feedback. This research was partially supported by a grant to the first author from the National Science Foundation (BCS-0819231).

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69*, 135–178.
Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.
Allen, C. (2009). Teleological notions in biology. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Winter 2009 Edition)*. <http://plato.stanford.edu/archives/win2009/entries/teleology-biology/>.
Atran, S. (1994). Core domains versus scientific theories: Evidence from systematics and Itza-Maya folkbiology. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 316–340). Cambridge, England: Cambridge University Press.
Atran, S. (1995). Causal constraints on categories. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 205–233). Oxford, England: Clarendon Press.
Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experiment Psychology: Learning, Memory, and Cognition, 11*, 629–654.
Barsalou, L. W., Sloman, S. A., & Chaigneau, S. E. (2005). The HIPE theory of function. In L. Carlson & E. van der Zee (Eds.), *Representing functional features for language and space: Insights from perception, categorization, and development* (pp. 131–147). Oxford, UK: Oxford University Press.
Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research, 18*, 433–447.
Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*, 1–29.
Bloom, P. (1998). Theories of artifact categorization. *Cognition, 66*, 87–93.
Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–185). Cambridge: MIT Press.
Carlson, G., & Pelletier, F. J. (Eds.). (1995). *The generic book*. Chicago, IL: University of Chicago Press.
Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science, 8*, 472–480.
Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition, 103*, 120–130.
Casler, K., & Kelemen, D. (2008). Developmental continuity in teleo-functional explanations: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development, 9*, 340–362.
Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General, 133*, 601–625.
Chaigneau, S. E., Castillo, R. D., & Martínez, L. (2008). Designers' intentions bias judgments of proper function, but they are not essences. *Cognition, 109*, 123–132.
Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*, 741–760.

Dawkins, R. (1995). *River out of Eden: A Darwinian view of life*. New York, NY: Basic Books.
Defeyter, M. A., & German, T. (2003). Acquiring an understanding of design: Evidence from children's insight problem solving. *Cognition, 89*, 133–155.
Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*, 164–168.
DiYanni, C., & Kelemen, D. (2005). Time to get a new mountain? The role of function in children's conceptions of natural kinds. *Cognition, 97*, 327–335.
Gelman, S. A. (2003). *The essential child: The origins of essentialism in everyday thought*. New York: Oxford University Press.
Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76*, 91–103.
German, T. P., & Johnson, S. (2002). Function and the origins of the design stance. *Journal of Cognition and Development, 3*, 279–300.
Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B, 205*, 581–598.
Greif, M. L., Kemler Nelson, D. G., Keil, F., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17*, 455–459.
Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441–461.
Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory & Cognition, 35*, 1785–1800.
Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Eds.). *Modularity and constraints in language and cognition: Minnesota symposium on child psychology* (Vol. 25, pp. 103–138). Hillsdale, NJ: Earlbaum.
Keil, F. C. (1994). The birth and nurturance concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). Cambridge, England: Cambridge University Press.
Keil, F. C. (1995). The growth of causal understanding of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multi-disciplinary debate* (pp. 234–262). Oxford, England: Clarendon Press.
Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Science, 3*, 461–468.
Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. In S. Laurence & E. Margolis (Eds.), *Creations of the mind: Theories of artifacts and their representation*. Oxford, UK: Oxford University Press.
Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*, 138–143.
Kemler-Nelson, D. G. et al (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development, 10*, 347–380.
Kemler-Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition, 77*, 133–168.
Kemler-Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*, 1271–1288.
Kim, S., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition, 39*, 649–665.
Kim, S., & Murphy, G. (2011). Ideals and category typicality. *Journal of Experimental Psychology: Learning Memory and Cognition, 37*, 1092–1112.
Laurence, S., & Margolis, E. (Eds.). (2007). *Creations of the mind: Theories of artifacts and their representation*. Oxford, UK: Oxford University press.
Leslie, S.J. (2011). Generics. In *Routledge Encyclopedia of Philosophy*.
Lin, E. L., & Murphy, G. L. (1997). The effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human and Perception Performance, 23*, 1153–1163.
Lombrozo, T. (2009). Explanation and categorization: How "why?" informs "what?". *Cognition, 110*, 248–253.
Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61*, 303–332.
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.
Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's Disease. *Psychological Science, 18*, 999–1006.
Lombrozo, T., Shtulman, A., & Weisberg, M. (2006). The Intelligent Design controversy: Lessons from psychology and education. *Trends in Cognitive Sciences, 10*, 56–57.
Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition, 28*, 41–50.
Malt, B. C., & Johnson, E. C. (1992). Do artifacts have cores? *Journal of Memory and Language, 31*, 195–217.
Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.
Matan, A., & Carey, S. (2001). Developmental changes within the core of artifact concepts. *Cognition, 78*, 1–26.
Medin, D. L., & Atran, S. (Eds.). (1999). *Folk biology*. Cambridge, MA: The MIT Press.
Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). Cambridge, MA: Cambridge University Press.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.
Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberchien (Ed.), *Advances in cognitive science. Theory and applications* (Vol. 2, pp. 23–45). Chichester, England: Ellis Horwood.
Nosofsky, R. M. (1988). Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning Memory and Cognition, 14*, 54–65.
Oakes, L. M., & Madole, K. L. (2008). Function revisited: How infants construe functional features in their representation of objects. In R. Kail (Ed.). *Advances in child development and behavior* (Vol. vol 36, pp. 135–185). San Diego: Elsevier.

Oppenheimer, D. M., Meyvisb, T., & Davidenkoc, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867–872.
Prasada, S., & Dillingham, E. M. (2006). Principles and statistical connections in common sense conception. *Cognition, 99*, 73–112.
Prasada, S., & Dillingham, E. M. (2009). Representations of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33*, 401–448.
Puebla-Ramirez, G. & Chaigneau, S. E. (2011). Is the centrality of design history function an effect of causal knowledge? In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1533–1538). Austin, TX: Cognitive Science Society.
Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). Oxford, UK: Oxford University Press.
Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 36*, 1171–1206.
Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition, 37*, 715–729.
Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 29*, 1141–1159.
Rehder, B. & Martin, J. (2011). A generative model of causal cycles. In L. Carlson. C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
Rips, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 21–59). Cambridge, England: Cambridge University press.
Rosch, E. H., & Mervis, C. B. (1975). Family resemblance. Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.
Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language, 37*, 240–267.
Shtulman, A. (2006). Qualitative differences between naïve and scientific conceptions of evolution. *Cognitive Psychology, 52*, 170–192.
Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189–228.
Strevens, M. (2000). The essentialist aspect of naïve theories. *Cognition, 74*, 149–175.
Truxaw, D., Krasnow, M. M., Woods, C., & German, T. P. (2006). Conditions under which function information attenuates name extension via shape. *Psychological Science, 17*, 367–371.
Wightman, C. (2008). *Sugar gliders* (2nd ed.). Hauppage, NY: Barron's Educational Series.
Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*, 449–468.
Wright, L. (1973). Functions. *The Philosophical Review, 82*, 139–168.