Cognitive Psychology 139 (2022) 101507

Contents lists available at ScienceDirect

Cognitive Psychology

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

How "is" shapes "ought" for folk-biological concepts

Emily Foster-Hanson *, Tania Lombrozo

Princeton University, United States

ARTICLE INFO

Keywords:
Concepts
Folk biology
Normativity
Causal reasoning
Functional explanation
Teleology

ABSTRACT

Knowing which features are frequent among a biological kind (e.g., that most zebras have stripes) shapes people's representations of what category members are like (e.g., that typical zebras have stripes) and normative judgments about what they ought to be like (e.g., that zebras should have stripes). In the current work, we ask if people's inclination to explain *why* features are frequent is a key mechanism through which what "is" shapes beliefs about what "ought" to be. Across four studies ($N = 591$), we find that frequent features are often explained by appeal to feature function (e.g., that stripes are for camouflage), that functional explanations in turn shape judgments of typicality, and that functional explanations and typicality both predict normative judgments that category members ought to have functional features. We also identify the causal assumptions that license inferences from feature frequency and function, as well as the nature of the normative inferences that are drawn: by specifying an instrumental goal (e.g., camouflage), functional explanations establish a basis for normative evaluation. These findings shed light on how and why our representations of how the natural world *is* shape our judgments of how it *ought* to be.

1. Introduction

> Zebras should have stripes, of course,
> So Zak thinks he should change.
> Liz Carter, *Zak the Zigzagged Zebra*.

Think of a zebra. Decades of research suggest that concepts are shaped by the statistical properties people experience (directly or indirectly), so the zebra that comes to mind will likely have black and white stripes, four legs, two eyes, and so on (Hampton, 1979; Nosofsky, 1988; Rosch & Mervis, 1975). These judgments of what is normal, or typical, reflect *descriptive* beliefs about the natural world. But representations of a "typical" zebra can also be shaped by *normative* beliefs or ideals: If you think that a zebra with stripes is a "better" zebra, you may judge it to be more typical and it might be more likely to come to mind (Barsalou, 1985; Bear et al., 2020; Bear & Knobe, 2017; Borkenau, 1990; Burnett et al., 2005; Foster-Hanson & Rhodes, 2019a; Lynch et al., 2000; Read et al., 1990; Rein et al., 2010). Concepts that encode both descriptive and normative components also have the potential to license normative judgments about category members: If striped zebras are better, you might agree with Liz Carter's poem in thinking zebras *should* have stripes.

These observations raise two related puzzles. First, where do normative judgments about how the natural world ought to be come from? (That is, why, and in what sense, *should* zebras have stripes?) And second, how and why do normative judgments relate to our representations of how the world *is*? (That is, why might a more *ideal* zebra be regarded as more *typical*?) In the current paper, we aim to solve these puzzles in the context of biological kinds.

* Corresponding author.
E-mail address: emily.fosterhanson@princeton.edu (E. Foster-Hanson).

https://doi.org/10.1016/j.cogpsych.2022.101507
Received 23 November 2021; Received in revised form 9 August 2022; Accepted 18 August 2022
Available online 13 November 2022

Both of these puzzles concern the relation between *is* (i.e., descriptive beliefs about what categories and their members are normally like) and *ought* (i.e., normative judgments of how categories or their members should be). As we discuss in greater detail below, prior work reveals that people sometimes reason that *what is* approximates or sheds light on *what ought to be* (Hume, 1739/2003; see also Black, 1964; Hudson, 1969; Knobe et al., 2013). This pattern of reasoning is sometimes referred to as the naturalistic fallacy (a term first used by Moore, 1903/2004, to describe is-ought reasoning in general; the fallacy of drawing such inferences from nature in particular was first described by Friedrich et al., 1989; Kierniesky & Sobus, 1989). This is a fallacy because, without the additional premise that the way things are is valuable or good, a normative claim does not follow from a descriptive one. For instance, the observation that most zebras *do* have stripes does not mean that zebras *should* have stripes, unless we have reason to think that the way zebras happen to be reflects something that is valuable or good – for example, that current organisms embody the ideals of some creator, or that natural selection results in biological features that are valuable and good.

In the current work, we propose and test a hypothesis about the missing premise from *is* to *ought* in laypeople's reasoning about biological kinds. The hypothesis is that this premise comes from people's folk biological beliefs about *why* certain features are prevalent. Specifically, we propose that people often posit *functional explanations* for the features of biological kinds (e.g., that zebras have stripes because stripes have served an important function, such as camouflage), reflecting their intuitive beliefs that the properties of biological kinds are typically the result of actual design (a creator) or apparent design (natural selection). Crucially, functional explanations bring with them a normative standard against which category members can be judged as better or worse (e.g., better or worse in terms of potential for camouflage; see Lombrozo & Wilkenfeld, 2019, for relevant discussion). This normative standard can define category ideals, shaping representations of what is typical, or "normal," for a category (Gelman & Legare, 2011; Medin et al., 1987; Murphy & Medin, 1985; Wattenmaker, 1999). This normative standard, and subsequent representations of what is "normal," can also license normative judgments about what category members *should* be like (an idea first posited by Aristotle, ca. B.C./1996; see also Lane, 2020). Finally, because people's folk-biological expectations about nature entail thinking that evolution means getting better over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014), these explanations also generate expectations that functional features will continue to become more frequent in the future. This proposal is summarized in Fig. 1.

In the remainder of the introduction, we first offer a brief review of prior work on frequency and ideals in shaping category representations (as reflected in judgments of typicality), as well as prior work on inferences from *is* to *ought*. We then develop our

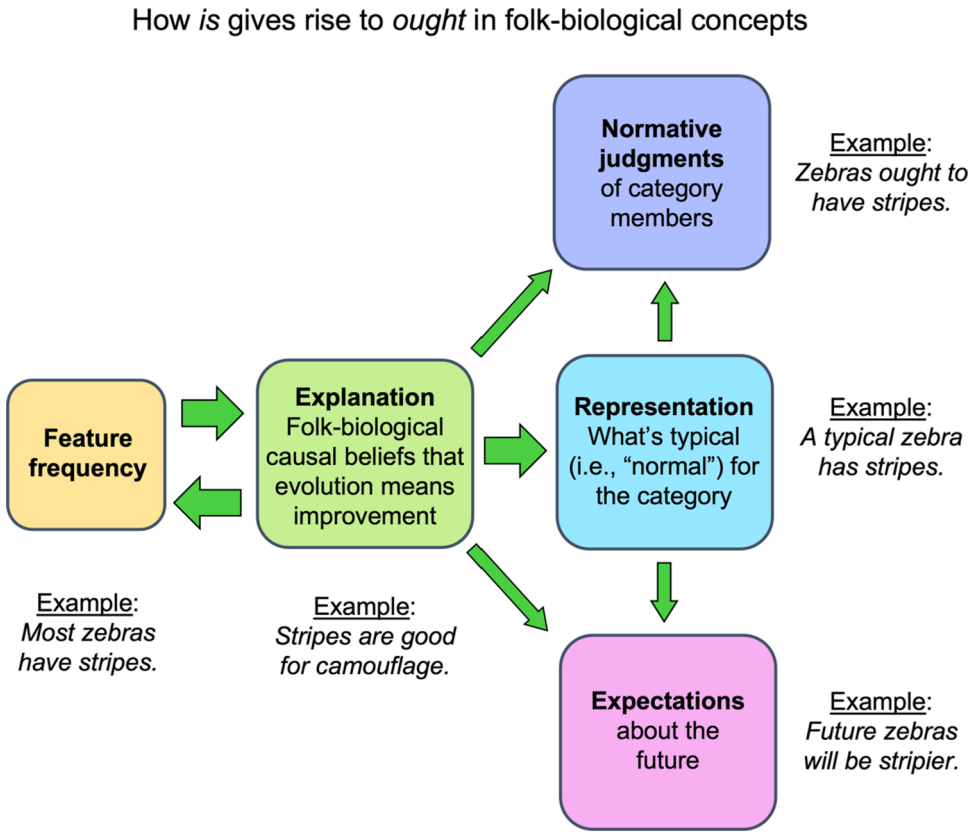

Fig. 1. Proposed causal process model. Arrows show the posited relationships that we tested across studies. This model includes two possible avenues through which descriptive information gives rise to normative judgments and expectations about the future: through explanations of feature frequency, or via category representations that encode what is "normal.".

proposal in greater detail and offer an overview of the four experiments we go on to report.

1.1. Frequency and category ideals shape typicality

People judge some category members as more typical than others (e.g., a robin is more typical than an emu for the category of birds), and this graded structure is one of the most robust phenomena in research on conceptual representation (e.g., Murphy, 2002; Rips, 1975; Rips et al., 1973; Rosch, 1973; Rosch & Mervis, 1975). Typicality has important consequences for how people use concepts to learn and reason in their daily lives: Typical category members come to mind more readily (Anglin, 1986; Rosner & Hayes, 1977), are often learned first (Bjorklund & Thompson, 1983; Rosch et al., 1976; Mervis & Pani, 1980; Rosch et al., 1976), and serve as the basis for people's inferences about what other category members are like (Osherson et al., 1990; Rips, 1975). This feature of concepts has often been referred to in the philosophical literature as "normality" (e.g., Bear & Knobe, 2017; Bear et al., 2020; McGrath, 2005; Wysocki, 2020), with what is "normal" (i.e., typical) determining what comes to mind when people think of a given category, and also serving as the standard against which individual category members can be compared (see Osherson et al., 1990).

People's graded category structure encodes their descriptive beliefs about feature frequency: Category members that are more typical tend to have more features in common with other category members, and fewer in common with non-members (i.e., family resemblance; Rosch & Mervis, 1975; see also Ameel & Storms, 2006; Davis & Love, 2010; Goldstone et al., 2003; Kim & Murphy, 2011; Medin et al., 1987). Indeed, one of the most well-established findings in research on conceptual representation is that frequent features are viewed as typical and diagnostic of category membership (Hampton, 1979; Nosofsky, 1988). However, feature frequency is not the sole determiner of typicality: People's *ideals* can also shape judgments of what is typical (Barsalou, 1985; Bear & Knobe, 2017; Borkenau, 1990; Burnett et al., 2005; Foster-Hanson & Rhodes, 2019a; Lynch et al., 2000; Read et al., 1990; Rein et al., 2010). For example, landscapers view taller, less weedy trees (i.e., trees that are more ideal for landscape design) as more typical (Lynch et al., 2000). More generally, category members are often seen as typical to the extent they approximate the category's "central tendency," but with an additional role for proximity to category ideals, which tend to be more extreme than average values (Barsalou, 1985). The notion that people's concepts are shaped by ideals is thus supported by decades of research, but has much older roots – more than 2,000 years ago, Plato proposed that humans understand all things in the world in reference to their most idealized forms (ca. 380 B.C./1974; see also Mohr, 1977; Ziff, 1972).

Although most research on the role of ideals in shaping typicality has focused on *anthropocentric* ideals – i.e., ideals relative to human goals – people also sometimes reason about biological kinds in terms of *biocentric* ideals – i.e., ideals relative to an animal's or species' own goals, grounded in beliefs about adaptive fitness (Foster-Hanson & Rhodes, 2019a; Foster-Hanson et al., 2020). For example, learning that a certain feature (e.g., a longer snout) is important for an animal's adaptive fitness goals (e.g., getting food to eat) leads people to view more idealized features (e.g., longer snouts) as more typical for the animal (Foster-Hanson & Rhodes, 2019a). Similarly, features that serve biological functions (vs those that do not) play a larger role in people's intuitive biological classifications (Lombrozo & Rehder, 2012), suggesting that organisms that better instantiate a kind's biocentric ideals are more likely to be judged as members of that kind.

While it's well-established that frequency and ideals both shape typicality, it is less clear how they interact, and why typicality might be a function of both. In other words, why would people's concepts encode a graded structure with both descriptive and normative dimensions? In the context of social categories, Del Pinal and Reuter (2017) propose that ideals are important for representations of some kinds of categories (e.g., scientists) because they reflect commitments to category-specific values (e.g., revising beliefs in light of evidence) that can predict future behavior and category membership (see also Foster-Hanson & Rhodes, 2019b; Knobe et al., 2013). More generally, descriptive information (about frequency) could be valuable in supporting inferences about how categories are *now*, while normative information could be useful in shaping judgments about the direction in which category members might change in the future. Relatedly, while differentiating between descriptive and normative information could be important for optimizing prediction (anticipating what is likely) versus intervention (bringing about what is valuable), typicality could be a compromise between both types of category information (Bear & Knobe, 2017; Bear et al., 2020; Wysocki, 2020), or even shift depending on the inferential goals relevant at the time of judgments (e.g., Barsalou, 1985; see also Vasilyeva et al., 2017; but see Kim & Murphy, 2011, for evidence of inflexibility in typicality judgments). In the current studies, we look at the relation between feature frequency, normative claims, and typicality, and we probe judgments about both the present and the future.

1.2. "Is" to "ought" reasoning

A separate body of research has documented how frequency can also license normative judgments, with people often reasoning that *what is* approximates or sheds light on *what ought to be* (Hume, 1739/2003; see also Black, 1964; Hudson, 1969; Knobe et al., 2013). This tendency has been given a variety of names (e.g., "the is-ought problem," "the naturalistic fallacy," "the existence bias," "the status quo bias"), but all are characterized by a tendency to assume that what is descriptively true is also good. Particularly relevant to the present research, people sometimes assume that what is *natural* must be right and good (e.g., Friedrich et al., 1989; Kierniesky & Sobus, 1989; Ismail et al., 2012). Yet similar tendencies to infer "ought" from "is" permeate people's judgments across domains, ranging from how many credits are the best requirement for a major (Eidelman et al., 2009) to judgments about the justness of social hierarchies (Kay et al., 2009; see also Jost & Banaji, 1994; Jost et al., 2004). Even young children make these inferences, often even more reliably than adults do (Foster-Hanson et al., 2021; Roberts et al., 2017; Tworek & Cimpian, 2016).

Yet people do not uniformly expect that what is frequent is good. Both children and adults think that category members should display frequent features that are central to category membership (e.g., judging that dogs *should* bark, and that there is probably

something wrong with one that does not), but they do not hold these same expectations for frequent features that are more peripheral (e.g., wearing collars; Haward et al., 2018; Prasada & Dillingham, 2006; 2009). Tworek and Cimpian (2016) propose that people's explanations about frequent features – rather than frequency per se – license these normative judgments. They suggest that people are inclined to explain frequent features as due to inherent causes, and that people tend to view inherent features normatively. In support of this proposal, they found that people who explain a phenomenon (e.g., giving roses on Valentine's Day) as due to inherent causes (e.g., roses' color or smell) rather than external causes (e.g., the cost of importing roses in February) are more likely to judge that the practice is right and good. Relatedly, Haward et al. (2021) argue that explaining features by appeal to category membership (e.g., dogs bark *because* they are dogs) indicates that the category and feature share a "principled connection," which in turn licenses the judgment that a category member without that feature has something wrong with it. Like these proposals, we suggest that people's explanations for frequent features are central to is-ought judgments. But instead of pointing to inherent or category-based explanations, we consider the role of people's causal explanations about why features became frequent to begin with – explanations that are often grounded in function.

1.3. A missing link: Functional explanation

When reasoning about the natural world, people's explanations are often characterized by an early-emerging and persistent tendency to assume that features are functional (Atran, 1994; Keil, 1994; Kelemen & Rosset, 2009; Kelemen et al., 2013; Medin & Atran, 1999). This tendency is particularly pronounced in early childhood (Kelemen, 1999; 2004) and in adults with limited formal education or impaired cognitive abilities (Atran, 1994; Lombrozo et al., 2007; Sánchez Tapia et al., 2016), suggesting that appeals to feature function may be an intuitive way in which humans explain the natural world. People also view functional features as central for category judgments about biological categories, and they expect functions to be stable and persistent across time (Lombrozo & Rehder, 2012).

That said, functional explanations are also highly selective, and not all functional explanations are appropriate, or equally good. For example, while many adults will agree that kangaroos have long tails for balance, most will deny that people have noses to hold up their glasses (even though noses serve this function; Liquin & Lombrozo, 2018). Lombrozo and Carey (2006) argue that functional explanations are a kind of causal explanation that is only licensed when two conditions obtain: (i) The function played a causal role in bringing about the feature being explained, and (ii) the process by which the function played a causal role must be generalizable within a predictable pattern. While these conditions are rarely observed to hold directly, they can be inferred based on more accessible properties such as structure–function fit (e.g., how well a given feature supports a candidate function, Liquin & Lombrozo, 2018), as well as intuitive theories of the natural world. In a domain like biology, assuming that a feature resulted from natural selection offers the conditions to meet these requirements (see also Wright, 1976). Moreover, common misconceptions about natural selection (e.g., that it is directed towards a goal, Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014) might make functional explanations of biological traits even more pervasive, and give rise to the assumption that future change will lead to greater functional optimization.

If our account is correct, then the missing premise from "is" to "ought" comes from the intuitive theories and causal commitments that lead people to posit functional explanations in response to salient patterns in the biological world (see Fig. 1). That is, people's beliefs that biological patterns reflect the outcome of a historical process directed towards what is functionally valuable or good then rationally support the explanatory belief that the pattern exists *because* it is functionally valuable or good. And because people encode these beliefs in their representations of what is typical, or normal, for a category, they offer a normative standard for evaluation of individual category members, as well as a basis for predicting category change. For example, if people think that a particular pattern is common (e.g., a zebra's stripes), then they may be inclined to assume that it serves a natural function, based on their beliefs about the underlying causal processes that drive changes in nature (e.g., camouflage from predators). These functional explanations would in turn generate evaluations about what is right and good, as described by the naturalistic fallacy (e.g., judging that zebras *should* have stripes).

Consistent with this proposal, recent work has found that people sometimes think of humans or entire organisms as having a particular function (e.g., to reproduce) and judge individuals who do not fulfill these functions (e.g., who choose not to reproduce) as immoral (Lewry et al., 2021). However, research to date has not investigated the more general hypothesis that functional explanations shape representations of biological kinds, with implications for typicality, normative judgments, and expectations about the future, as depicted in Fig. 1.

1.4. The current studies

In the current studies, we test our causal model in two parts. First, we test the prediction that manipulating frequency affects people's judgments of both typicality and normativity (i.e., is-to-ought judgments), and that it does so in large part because frequent features encourage people to posit functional explanations for those features. In Study 1, we show that manipulating frequency affects functional explanations, typicality, and normative judgments as we would expect (though ceiling effects and poor explanations limit the potential to test all predicted effects). In Study 2, we address these concerns, and moreover show that we can control the effect of frequency on category judgments by manipulating the causal process that gives rise to frequent features. According to our model, functional explanations are only licensed when they are causal: it must be the case that the functional feature resulted from a consequence-driven process through which a feature's function *caused* it to exist or become frequent. When features do not support these causal assumptions, people should therefore be less inclined to offer functional explanations. We find evidence in support of this

prediction in Study 2.

Studies 1 and 2 manipulate functional explanations indirectly (by manipulating frequency and context), and thus the relation between feature function and people's category judgments is correlational. In our final two studies, we show causal effects of function directly through an experimental manipulation of functional explanations. In Study 3, we show that manipulating functional explanations affects people's predictions about feature frequency, as well as their typicality and normative judgments, as our model would predict. In Study 4, as in Study 2, we show that these effects are moderated by the causal history of functional features. We also test whether functional explanations are particularly important for people's judgments about how categories are likely to change in the future. We find preliminary support for this prediction in Study 2, and stronger support for this prediction in Study 4.

2. Study 1

Study 1 had three central aims. Our first aim was to test the hypothesis that salient patterns in the natural world (in this case, uneven distributions of features) call out for explanation, and do so in such a way that the most frequent features of biological organisms are (often) explained by appeal to a function. Our second aim was to test the hypothesis that frequent features are more likely to be regarded as highly typical (consistent with prior work), but also more likely to support normative claims about how an organism ought or ought not to be (going beyond prior work). Finally, our third aim was to test the hypothesis that functional explanations predict judgments of typicality and normative claims, consistent with the hypothesis that frequency influences typicality and normative claims in part because it prompts functional explanation.

To test these hypotheses, we presented participants with novel biological organisms that varied with respect to feature prevalence. We then elicited explanations, typicality judgments, and normative evaluations. Study 1's design and analyses were preregistered on OSF, https://osf.io/j374z.

2.1. Methods and materials

2.1.1. Participants

As specified in our preregistration plans, our desired sample size across all four studies was 24 participants per condition, based on previous work using these materials (Foster-Hanson & Rhodes, 2019a). In Study 1, our desired sample was therefore 96 participants, so we recruited 106 participants to allow for possible drops. Participants were recruited using Prolific and tested using Qualtrics; they received $1.50 for participating. Although we indicated on Prolific that eligible participants should be located in the United States, 9 participants who completed the study had an IP address that identified them as being outside of the United States and were therefore excluded. All remaining participants correctly answered at least 8 out of 11 attention and manipulation check questions throughout the study, so we retained the full eligible sample of 97 participants for analysis, per our preregistration (56 female, 40 male, 1 non-binary; $M_{age} = 35$). Study procedures for all studies were approved by the Institutional Review Board of the authors' university.

2.1.2. Procedure

Participants in all four studies first completed a short typicality training to ensure they understood the typicality task that they would later perform. This training was based on that used in previous work (Foster-Hanson & Rhodes, 2019a; Kim & Murphy, 2011; Rosch & Mervis, 1975). After the training, participants learned about four novel animals, each presented visually with five category members that varied along a single feature dimension, such as number of spots (Foster-Hanson & Rhodes, 2019a; see Fig. 2). On each trial, participants learned that the most common feature value was either on the far left of the scale (skewed-left, two trials) or the far right of the scale (skewed-right, two trials; counterbalanced within and between participants using a Latin square design). After each animal was introduced, participants answered explanation and typicality questions about that animal, detailed below, before moving on to the next animal.

To test the prediction that people are inclined to explain high-frequency features by appealing to their function, we asked participants to provide an explanation for why the animal's features were distributed as they were, which we coded for references to feature function. To assess typicality, we asked which category member participants viewed as most representative (i.e., "Which X is the most typical X?") and informative (i.e., "Which X would you look at to learn the most about Xs?"). Question order was counterbalanced between-participants, with each participant completing either typicality questions first or explanation questions first.

After answering the explanation and typicality questions for all four animals, participants evaluated normative claims and answered forced-choice function questions. For the normative judgments, participants rated their agreement with normative *ought* claims about how each animal ought to be on a 7-point Likert scale (1 = completely disagree, 7 = completely agree). These claims described both frequent features (e.g., "Virdexes ought to have very many spots") and infrequent features (e.g., "Virdexes ought to have very few spots"), presented in counterbalanced order within-participants.[1] Participants also answered a forced-choice function question, reporting whether they: (a) had thought that more frequent features were better for each animal, (b) had thought that less frequent features were better (presented in random order), or (c) did not think at all about which features were better. These normative judgments and forced-choice function questions were included after the four animal trials to ensure that if participants showed a tendency towards functional reasoning or an influence of normative beliefs in their explanations and typicality judgments, it was not

[1] For normative questions about the last animal trial only, half of participants saw a picture showing the opposite distribution than the one they had been assigned to by mistake; these responses were excluded from analysis.

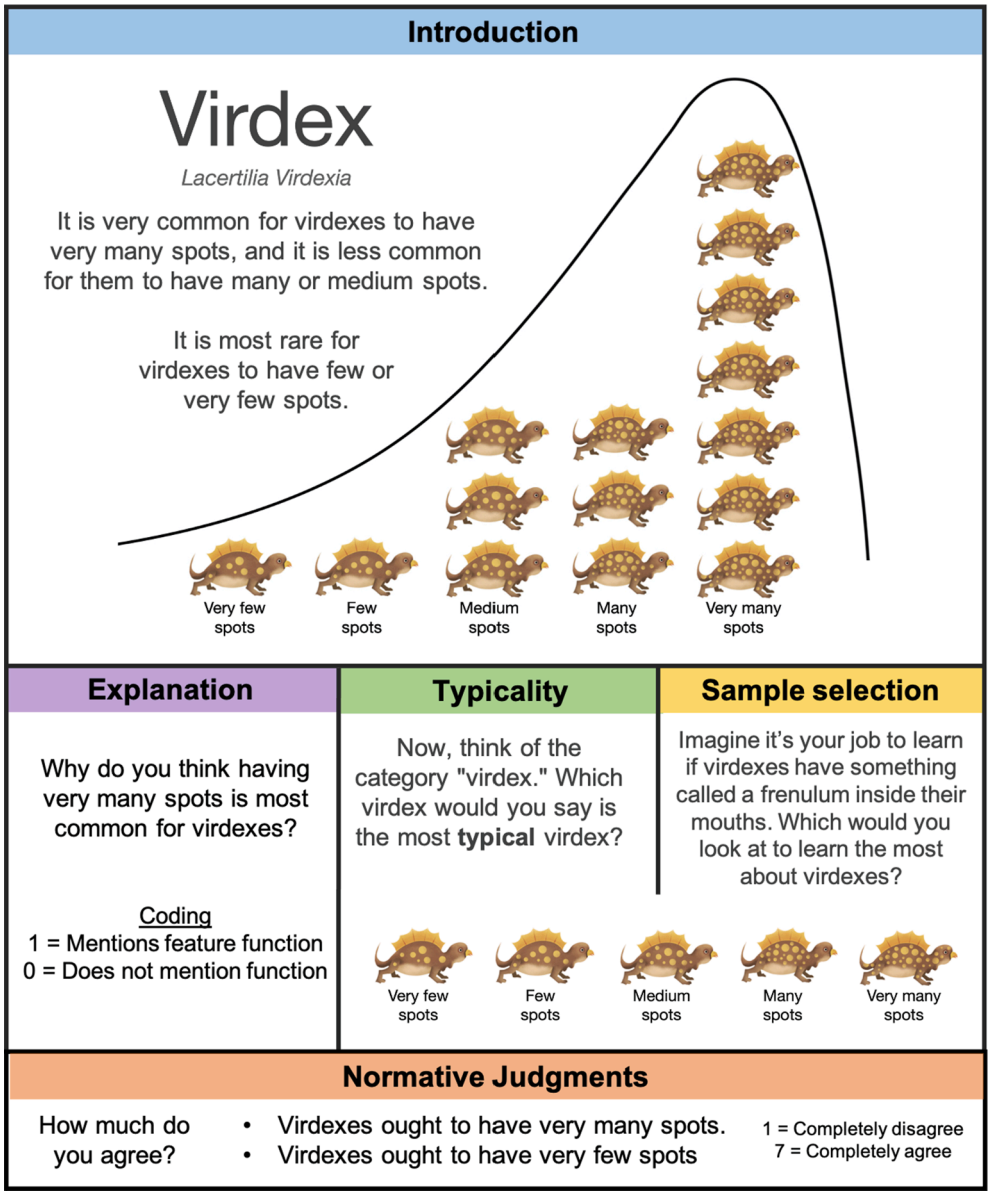

Fig. 2. Method, Study 1. On each animal trial, participants learned about a novel animal kind from five category members varying along a single feature dimension. They learned that some feature values were more frequent than others. Then they gave an explanation for the feature frequency and answered questions about typicality, in counterbalanced order (between participants). On typicality measures, participants responded by clicking on the image corresponding to their choice. The entire procedure then repeated for the remaining three animal kinds. After completing all four animal trials, participants evaluated normative claims and answered forced-choice function questions.

because the study itself made these dimensions salient. Study materials and procedures can be found on OSF, https://osf.io/863rc/.

At the end of the study, we also administered a battery to assess understanding of natural selection, as well as a measure of acceptance of natural selection, using a set of questions from Weisberg et al., (2021; 8 questions), and Gaia beliefs (measured as agreement with the statement "nature is a powerful being"). Most participants (55%) reported having naturalistic beliefs about the origins of the natural world (believing that animals and plants developed entirely through natural processes); 13% reported having creationist beliefs (animals and plants were created by God in more or less their current form), 12% reported theist beliefs (animals and plants developed through natural processes, which were guided by God the entire time), and 20% reported deist beliefs (animals and plants developed through natural processes, which were set up by God but continued on their own). Participants' responses on our variables of interest did not vary depending on their beliefs about nature or understanding of evolution (all *ps* greater than .05), therefore we do not consider these questions further. Data and analysis code for all four studies are available on OSF, https://osf.io/863rc/.

2.2. Results

2.2.1. Do people explain frequency by appealing to feature function?

We coded participants' open-ended explanations about *why* certain features were frequent for whether they referenced function (coded as 1) or not (coded as 0). Two independent raters coded each explanation (initial agreement 97%; Cohen's Kappa = 0.94; disagreements resolved through discussion). Roughly half of participants (45%) explicitly described feature function in their explanations (e.g., "better camouflage"), and only 1 of these explanations described functions unrelated to frequent features, compared to 171 that described frequent features as functional. This difference was highly significant (X^2 (1) = 39.15, $p < .001$). These findings support the hypothesis that uneven distributions of features are often explained by appeal to the function of prevalent features.

Participants' responses to forced-choice function questions (which asked them at the end of the task to report whether they thought features were functional) mirrored these results. Responses varied by distribution (X^2 (1) = 18.31, $p < .001$), with participants most often reporting that they thought frequent features were better for the animal. For exploratory purposes, we re-coded responses to forced-choice questions for whether the most common feature was better for the animal (1 = thought common feature was better, 0 = all other responses). In these analyses, participants reported thinking that the most frequent features were better for the animals at above chance levels (61% of trials, comparison to chance, $p = .04$). Participants' forced-choice responses were also positively correlated with their coded open-ended explanations ($r(386) = 0.52$, $p < .001$). Again, this suggests that many participants were spontaneously inclined to explain unequal distributions of features by assuming that common features were functional.

2.2.2. Do feature frequency and functional explanations shape typicality?

We analyzed participants' typicality and sample selection responses using ordinal logistic regression mixed models (CLMMs; Christensen, 2019), with distribution condition as a predictor. We included random intercepts for order, participants, and trials, and we report the results of Likelihood Ratio Tests. Participants viewed category members with the most frequent features as most typical (X^2 (1) = 444.14, $p < .001$) and informative of their kinds (X^2 (1) = 392.52, $p < .001$). Individual participants' typicality and sample selection responses were highly correlated ($r(386) = 0.93$, $p < .001$). As described in our preregistration plan, we re-analyzed participants' typicality and sample selection responses including references to function as a predictor variable and testing for main and interactive effects.[2] Participants who referred to the function of the more prevalent feature to explain the feature distribution also picked more extreme category members as typical (X^2 (1) = 6.89, $p = .009$) and informative of their kinds (X^2 (1) = 6.43, $p = .01$; Fig. 3). Patterns were similar when including forced-choice function attributions as a predictor of typicality responses (X^2 (1) = 4.79, $p = .03$), although this effect was not significant for judgments of informativeness (X^2 (1) = 2.20, $p = .14$).

2.2.3. Do feature frequency and functional explanation shape normative evaluations?

We analyzed agreement with normative claims using linear mixed models (LMMs; Kuznetsova et al., 2017), with distribution condition and claim type (claims about frequent features or infrequent features) as predictors, testing for main and interactive effects. We included random intercepts for order, participants, and trials, and we report the results of Likelihood Ratio Tests. Participants generally agreed with claims that animals ought to have frequent features and disagreed with claims that animals ought to have infrequent features (main effect of claim type, X^2 (1) = 1666.83, $p < .001$; Fig. 3). Thus, people think that category members *ought* to display common features. When we repeated the analysis for normative judgments including the provision of a functional explanation as an additional predictor, we did not find interactive effects of spontaneous functional explanations (X^2 (1) = 0.11, $p = .74$) or forced-choice function attributions (X^2 (1) = 2.92, $p = .09$). However, exploratory analyses including participants' typicality responses as a predictor revealed that participants who held more extreme representations of what was typical for the animals agreed more strongly with normative claims that they ought to have frequent features (three-way distribution × typicality × claim type interaction, X^2 (1) = 4.31, $p = .04$).

2.3. Discussion

When asked to explain why the prevalence of a set of features varied across a biological population, roughly half of participants explicitly assumed that common features serve an important function for the organism. Participants also judged that category members with more frequent features were more typical, and they indicated that category members *ought* to have frequent features (and ought *not* to have infrequent ones). These findings confirm our first two predictions: People often explain frequency by assuming that common features are functional, and they use frequency to predict both what members of biological categories typically *are* like and what they *should* be like. Our third prediction concerned the role of functional explanation in shaping typicality and normative judgments. For this prediction, our evidence was mixed. On the one hand, participants who referred to feature function to explain frequency chose more extreme category members as typical and informative of their kinds than those who did not mention feature function. On the other hand, although participants overwhelmingly agreed with normative claims that matched frequency and disagreed with opposite claims, we did not find an effect of functional explanations on these judgments. We did, however, find an effect of

[2] In exploratory analyses of the effect of question order, participants' sample selection (but not typicality) responses were also more extreme when they answered typicality and sample selections before being asked to explain the distribution of features (two-way question order x distribution interaction: X^2 (1) = 5.53, p =.02).

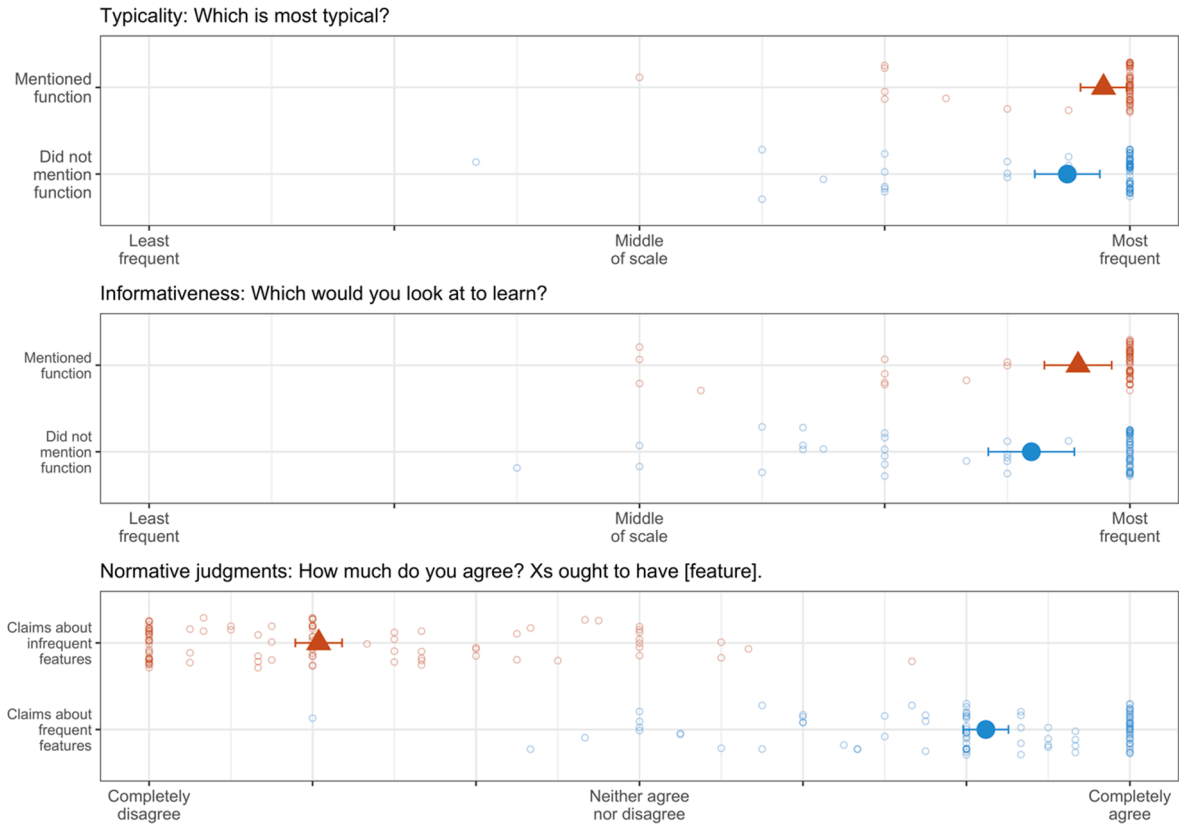

Fig. 3. Participants' responses in Study 1, collapsed across trials. The top and center panels show typicality and sample selection responses, by whether participants referred to feature function to explain the distribution. Responses on these measures are reverse-coded for the skewed-left distribution trials so that the most frequent category member is to the right for all trials. The bottom panel shows participants' agreement with normative claims that category members ought to display frequent or infrequent features. For all panels, large shapes show group means with 95% Confidence Intervals, and small circles are individual participant averages.

typicality on normative judgments in exploratory analyses, suggesting that both explaining feature frequency in terms of functions, and encoding these explanations in representations of what is typical or "normal" for the category, may work together to license judgments of what individual category members ought to be like.

The data in Study 1 could have been inconclusive with respect to our third hypothesis for two reasons. First, some participants' responses were ambiguous or too short to code for explicit references to function (e.g., many participants simply wrote "evolution" or "genes"). Given these ambiguous responses, it is possible that a greater proportion of participants in fact assumed that frequent features were functional than we were able to detect. Second, because participants showed such enormous effects of feature frequency, we were limited in our ability to detect more subtle influences at the extremes of the scale. Finally, more robust support for the hypothesis that frequency influences typicality and normative claims in part because it prompts functional explanation would come from evidence of mediation. Unfortunately, this was not possible given Study 1's design because we did not manipulate whether participants observed a more frequent feature (all participants did), nor did we manipulate the link between frequency and functional explanation. Study 2 addressed these limitations.

3. Study 2

In Study 2, we had two main goals. Our first goal was to provide a stronger test of our central hypothesis by addressing the shortcomings in Study 1's design, including participants' short explanations and potential ceiling/floor effects. In addition, our hypothesis is that frequency shapes category beliefs *via* people's functional explanations, but testing for mediation would require an experimental manipulation of the conditions that prompt functional explanations. To accomplish this, we introduced a "context" manipulation. In the natural context, which mimicked the conditions from Study 1, participants were free to assume that more frequent features resulted from natural processes, and so we would expect their reasoning to reflect their beliefs about evolution and natural selection, particularly the expectation that evolution means getting better over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014). In the accidental context, we blocked these inferences from frequency by stipulating that the current distribution was the result of temporary accidental factors rather than stable natural processes. By varying whether frequency licensed a functional explanation,

we could test whether the effect of frequency on typicality and normative judgments is mediated by people's functional explanations.

Our second goal in Study 2 was to more accurately describe the character of people's is-to-ought reasoning about biological kinds by measuring agreement with a broader set of normative claims. The literature on normativity is somewhat mixed: Although normative judgments have often been investigated relative to social agents with the free will to disobey norms (Darwall, 2013; Strawson, 1962) and as part of humans' unique ability to collaborate and coordinate social action (Tomasello, 2020), people sometimes apply normative judgments to a much broader range of categories than this account would predict (). How do the normative judgments that people apply to social agents with free will differ from the judgments they apply to (non-human) biological kinds? In the current work, we attempt to clarify this debate by taking cues from the philosophical literature and differentiating between different types of normative judgments. In Study 1, participants agreed that animals ought to display frequent features, but *ought* statements are ambiguous regarding several different kinds of claims (von Fintel & Iatridou, 2008; Sloman, 1970). For example, the statement "Sam ought to be in her office" can refer to teleological or instrumental norms, which assign value to actions or features relative to specific goals (e.g., Sam ought to be in her office because the action is aligned with her goal of submitting an important grant application). Modal *ought* statements can also refer to deontic norms, which describe obligations and restrictions (e.g., Sam ought to be in her office because she has an obligation to her students to be available). Finally, ought statements can simply express epistemic norms, describing what is likely given the available information (e.g., Sam ought to be in her office because she is usually there at this time of day).

In Study 2, we dissociate these different interpretations of ought statements by asking participants to rate their agreement not only with claims about which features animals ought to have, but also with claims about teleological and deontic norms. If participants in Study 1 agreed with normative claims because they interpreted them as teleological, then we expect participants in Study 2 to show similar patterns of agreement for *ought* statements and for statements about teleological norms (e.g., asking which features it would be best for an animal to have). Functional reasoning establishes a normative standard against which category members can be judged as better or worse relative to specific functional goals. Thus, if functional explanations are a key mechanism through which information about frequency licenses normative judgments, then participants who appeal to function should agree with claims about teleological norms more strongly than participants who do not view common features as functional.

In contrast, if participants in Study 1 interpreted the *ought* statements as describing deontic obligations, then participants in Study 2 should agree with statements about deontic norms (e.g., asking whether animals with uncommon features have done something wrong) similarly to *ought* claims. Deontic judgments are closely linked to beliefs about agency and punishment, and logically should not apply to animals in the current studies who would have no control over their physical features (but see Goodwin & Benforado, 2015; Tabb et al., 2019). For these reasons, we expected agreement with these claims to be low overall. However, including these measures also allowed us to test whether deontic normative judgments might be sensitive to effects of context. If so, then this would suggest that different types of normative judgments are not as cleanly differentiated as one might expect (see Phillips et al., 2019., for relevant discussion of modal reasoning). We revisit this possibility in the General Discussion.

Finally, if participants merely interpreted the normative claims in Study 1 as epistemic (i.e., that animals would be predicted to have common features based on the available information), then participants in Study 2 should agree only with claims regarding which features animals "ought" to have, as in Study 1, but not with claims about teleological or deontic norms. This interpretation of the ought claims in Study 1 would be akin to a manipulation check—we told participants that certain features were more common, and they agreed that category members would therefore be most likely to display those common features.

As a final normative benchmark, we also asked participants to rate their agreement with judgments of nonconformity (i.e., whether there is something wrong with animals that do not have common features) to test which type of judgments align most closely with this measure of normativity that has been used in previous work (e.g., Haward et al., 2018; 2021; Prasada & Dillingham, 2006; 2009). Study 2′s design and analyses were preregistered on OSF, https://osf.io/jz6qh.

3.1. Methods and materials

3.1.1. Participants

Our desired sample size was again 24 participants per condition based on previous work using these materials, so we recruited 212 adults to allow for possible drops. Of these, 11 were excluded for having an IP address that identified them as being outside of the United States, and we excluded 4 additional participants for incorrectly answering more than 3 out of 14 attention and manipulation check questions throughout the study, per our pre-registration plan. This left 197 participants (111 female, 79 male, 6 non-binary; 1 who preferred to self-identify; $M_{age} = 36$). Participants were recruited using Prolific and tested using Qualtrics; they were paid $1.90.

3.1.2. Procedure

In this study, participants again learned about four novel animals and learned that some features were more common than others, as in Study 1. However, we varied the causal context that gave rise to the current distribution: Half of participants learned that all four animals lived in their natural habitat undisturbed by humans (natural context), while the other half learned that all four animals had been bred in captivity and that their current distribution was the result of temporary, accidental factors (accidental context; Fig. 4). We expected that participants who learned that the current distribution was the result of temporary circumstances would be less likely to reason that frequent features were functional, so information about frequency would therefore be less central to their category judgments.

We also suspected that functional reasoning might become particularly important when reasoning about how animals might change in the future. People expect functional features to be stable and persistent across time (Lombrozo & Rehder, 2012), and they

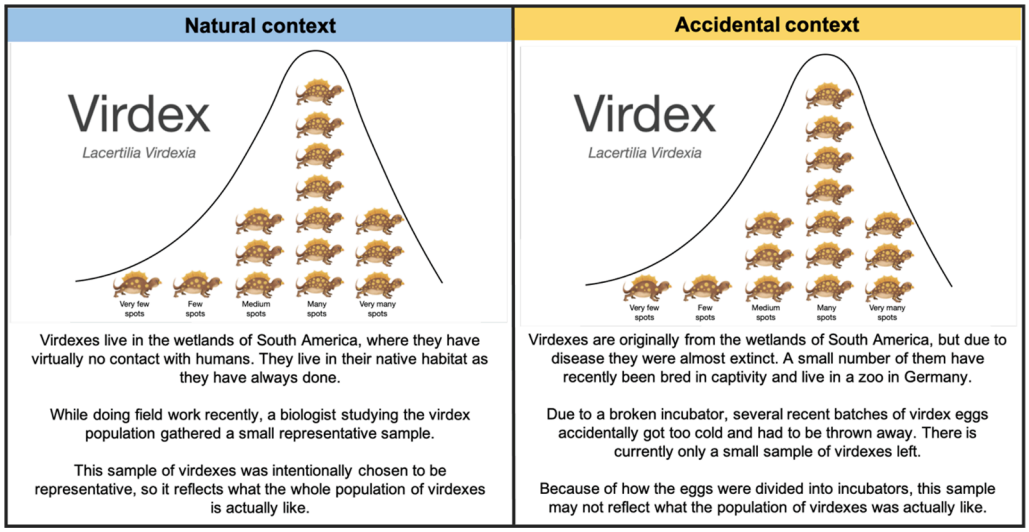

Fig. 4. Context manipulation, Study 2. Participants were randomly assigned to learn either that all four animals lived in their natural habitat undisturbed by humans (natural context), or that the current distribution of all four animals was the result of temporary, accidental factors (accidental context).

view stable causes as better explanations (Vasilyeva et al., 2018). However, people also tend to misunderstand evolution as goal-directed change in individuals, and to think of evolution as improvement over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014). For these reasons, we also sought to test whether functional information might be more influential when reasoning about the future than when reasoning about the present. To do so, we varied the time-period that participants were reasoning about within-participants: For the first two trials, all participants answered questions about the animals as they currently are, and for the third and fourth trials they made judgments about the animals 100 generations in the future (participants in the accidental context condition were told that the animals bred in captivity had been reintroduced into their natural habitat for these trials).

Distribution condition was within-participants, using a Latin square design: On each of four animal trials, participants learned that the most common feature value was either on the left of the scale (skewed-left, two animal trials) or the right of the scale (skewed-right, two animal trials; modified so that all participants saw one of each distribution in each time-period block). However, given the possibility of ceiling effects in Study 1, in Study 2 we showed participants the two next-most skewed distributions (i.e., with category members 2 or 4 on the scale as the most common, rather than 1 or 5 as in Study 1; see Fig. 4). This also allowed us to test whether participants might expect that the animal categories would become more skewed in the future than the present.

On each trial, after learning about the distribution of the animal's features, as well as the historical context that gave rise to that distribution, we asked participants whether they thought there was a reason for the current distribution of each animal's features or if it was simply random. We included this question to avoid inviting participants to reason about feature function if they would not otherwise have done so. We then asked participants to provide an open-ended explanation. Because many explanations provided by participants in Study 1 were ambiguous or too short to code for references to function, in Study 2 we requested that participants be as specific as possible and write at least one full sentence (40 characters). For participants who answered that there was a reason for the distribution, we coded explanations for what that reason was as in Study 1 (1 = referred to frequent feature function, 0 = did not refer to frequent feature function). One independent rater coded each explanation, and a second rater coded a randomly selected 20% of explanations for reliability (initial agreement 99%; Cohen's Kappa = 0.98; disagreements resolved through discussion). For participants who answered that it was simply random, we asked why they thought it was random so that the procedure was similar for all participants.

As in Study 1, we then asked which category member participants viewed as most representative of its kind (i.e., "Which X is the most typical X?") and most informative (i.e., "Which X would you look at to learn the most about Xs?"). We also measured how participants thought the feature in question varied across category members (or would vary 100 generations in the future), by asking participants to choose one distribution of features (of five possible, presented in random left–right or right-left order) that they thought described each animal category. Each distribution showed one of the five category members on the scale as most frequent (see Fig. 9). Participants completed a short training task at the start of the study to ensure they understood the predicted distributions task. We had initially intended to counterbalance question order (distribution questions first, or typicality and sample selection first) as a between-participants factor, but due to experimenter error all participants were assigned to answer typicality and sample selection questions first, followed by distribution questions. Note that in Study 1 this order of questions led to (slightly) less skewed responses on all three measures.

After concluding the four animal trials, participants answered questions probing their normative judgments about the animals, as in Study 1. However, in addition to indicating their agreement with overall normative claims that the animals ought to have frequent and

infrequent features (e.g. "Xs ought to have [feature]"), we also included three additional statement pairs to examine the nature of these normative judgments. These included claims about teleological norms (e.g., "It would be best for an X to have [feature]"), negative judgments of nonconformity (e.g., "If an X doesn't have [feature], there's probably something wrong with it"), and deontic norms (e.g., "If an X doesn't have [feature], then it's done something wrong").

3.2. Results

3.2.1. Do functional explanations vary by frequency and context?

Most participants (90%) who learned that the animals lived in their natural habitat said that there was a reason for the current distribution of each animal's features, compared to fewer than half (48%) of those who learned that the current distribution was due to accidental factors (main effect of context: X^2 (1) = 53.17, $p < .001$). These responses also interacted with time (context × time interaction: X^2 (1) = 6.31, $p = .01$), with participants in the natural context condition saying that there was a reason more often in the future ($M = 0.92$, 95% CI [0.83, 0.96]) than in the present ($M = 0.87$, 95% CI [0.76, 0.94]), but the reverse pattern for those in the accidental context condition (future: $M = 0.41$, 95% CI [0.25, 0.60]; present: $M = 0.54$, 95% CI [0.36, 0.72]).

When we analyzed participants' coded open-ended explanations for what that reason was, those in the natural context condition were significantly more likely than those in the accidental context to describe the function of common features (e.g., "I think that it's useful for camouflage"; main effect of context: X^2 (1) = 31.48, $p < .001$; Fig. 5). In contrast, participants in the accidental condition often simply restated the accidental cause that had led to the current distribution as the reason (e.g., "The egg incubator broke causing some eggs to be lost"). These references to function were also more common overall for the future than for the present (main effect of time: X^2 (1) = 4.07, $p = .04$). There were no other main nor interactive effects on this measure.

3.2.2. Do frequency and context shape typicality?

Typicality judgments varied by context: Participants in the natural context condition generally chose the most frequent category member as typical (i.e., 2 for the skewed left distribution, $M = 2.03$, 95% CI [1.94, 2.12]; 4 for the skewed right distribution, $M = 3.97$, 95% CI [3.89, 4.06]) but participants in the accidental condition chose category members closer to the middle of the scale as most typical (i.e., 3; skewed left: $M = 2.28$, 95% CI [2.17, 2.39]; skewed right: $M = 3.73$, 95% CI [3.62, 3.83]; two-way context × distribution interaction, X^2 (1) = 34.59, $p < .001$; subsumed main effect of distribution, X^2 (1) = 645.50, $p < .001$). Choices of which category members were most informative showed the same pattern as typicality judgments (context × distribution interaction: X^2 (1) = 26.93, $p < .001$; subsumed main effect of distribution: X^2 (1) = 534.37, $p < .001$). Participants in the natural context condition again generally chose to learn from the most frequent category member (skewed left: $M = 2.09$, 95% CI [1.99, 2.18]; skewed right: $M = 3.94$, 95% CI [3.85, 4.03]) whereas participants in the accidental condition chose to learn from category members closer to the middle of the scale (skewed left: $M = 2.35$, 95% CI [2.24, 2.47]; skewed right: $M = 3.72$, 95% CI [3.61, 3.83]). There were no main nor interactive

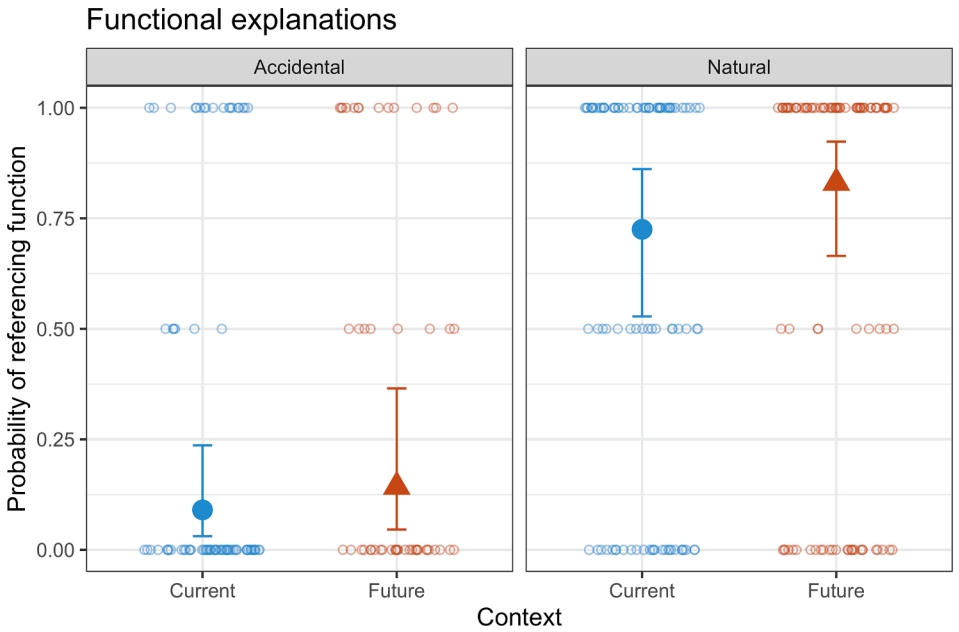

Fig. 5. Study 2: References to function in open-ended responses for participants who indicated that there is a reason for current distribution of features (90% of those in the Natural Context, and 48% of those in the Accidental Context). Responses are indicated by time period (current and future, within participants), collapsed across trials. Panels show the two context conditions (accidental and natural, between participants). Large shapes show group means with 95% Confidence Intervals, small circles are individual participant averages.

effects of time on either measure.

Distribution responses showed a three-way context × time × distribution interaction (X^2 (1) = 5.27, p = .02; subsumed context × distribution interaction, X^2 (1) = 16.57, p < .001; subsumed main effect of distribution: X^2 (1) = 629.43, p < .001). Participants chose similar feature distributions in the present across contexts (pairwise contrast, skewed left: p = .93; skewed right: p = .90), but they expected future distributions to become more skewed in the natural context, and more normal in the accidental context (pairwise contrast, skewed left: p = .02; skewed right: p = .007).

Individual participants' responses were highly correlated across typicality, sample selection, and distribution measures (all ps < .001).

3.2.3. *Do frequency and context shape typicality via functional explanation?*

As part of our preregistered analyses, we re-analyzed participants' responses including whether or not they generated a functional explanation as a predictor variable and testing for main and interactive effects. In these analyses, participants who referenced feature function to explain the current distribution of an animal's features chose more skewed category members as typical (two-way interaction between function reference and distribution: X^2 (1) = 11.43, p < .001; Fig. 6; two-way context × distribution interaction: X^2 (1) = 17.78, p < .001). Sample selection responses again showed the same pattern as typicality, with participants who referenced feature function to explain the current distribution of an animal's features choosing to learn from more skewed category members than those who did not reference function (two-way function reference × distribution interaction: X^2 (1) = 10.42, p = .001; two-way context × distribution interaction: X^2 (1) = 12.82, p < .001; Fig. 6). There were no main nor interactive effects of time on either measure. Distribution responses also showed this pattern (two-way function reference × distribution interaction (X^2 (1) = 4.21, p = .04; Fig. 6), as well as a three-way context × time × distribution interaction (X^2 (1) = 4.92, p = .03).

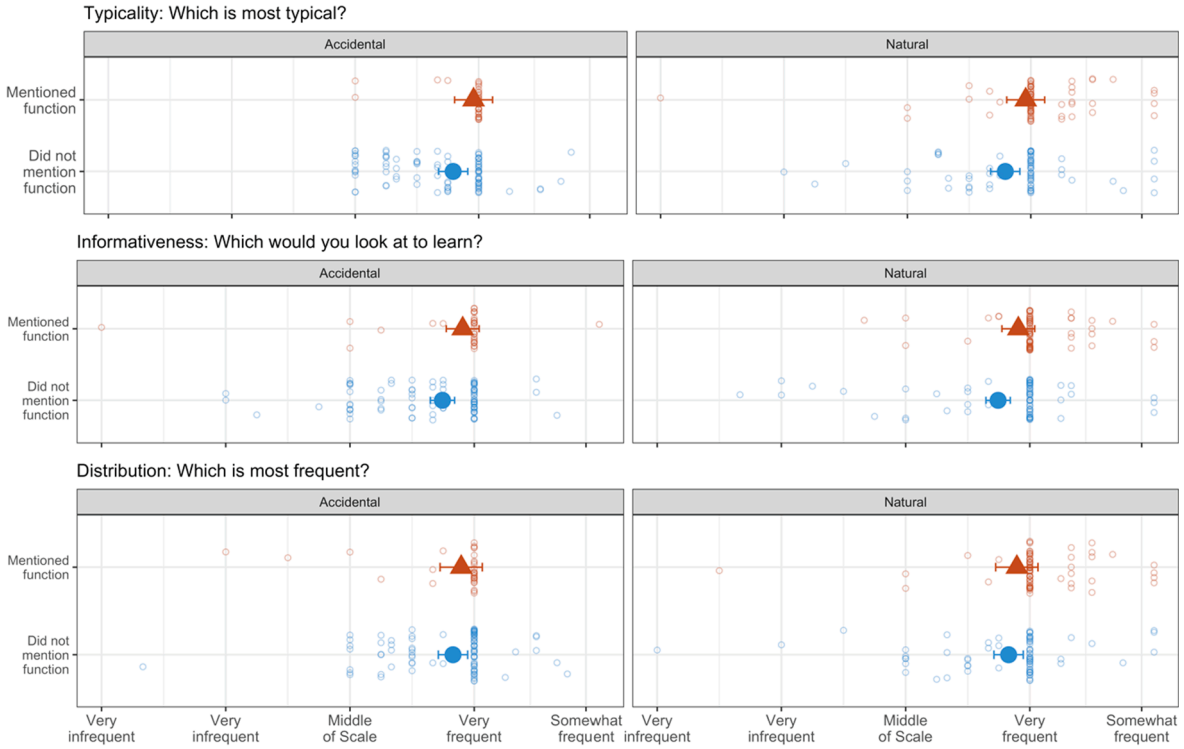

Fig. 6. Study 2: Participants' typicality, sample selection, and distribution responses, collapsed across trials, by whether or not participants referred to feature function to explain the distribution. Responses for the skewed left distribution trials are reverse coded so that the most frequent category member is in the same location on the y-axis (i.e., 4) for all trials. Left-right panels show the two context conditions (natural and accidental, between participants). Large shapes show group means with 95% Confidence Intervals, small circles are individual participant averages. Participants indicated their responses to typicality and sample selection measures by clicking on one animal in each 5-item scale; for distribution responses they clicked on one of five pictures showing different feature distributions. Note that 78% of explanations referenced function among participants in the natural context, compared to 11% of explanations among participants in the accidental context.

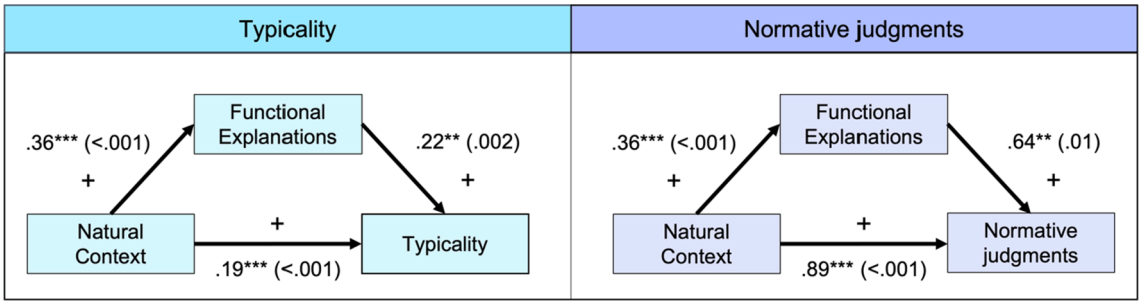

Fig. 7. Mediation analysis of participants' typicality and normative judgment responses in Study 2. The effect of context on typicality and normative judgments was incompletely mediated by whether or not participants generated a functional explanation.

When analyzing the probability of predicting a change in the animal's distribution of features, with references to function included as a predictor, the effect of time interacted with context and function references (three-way interaction: X^2 (1) = 13.02, $p < .001$), but the *direction* of predicted change varied only by context and references to function (X^2 (1) = 16.19, $p < .001$). Most participants in the natural context condition who predicted change thought the distribution would become more skewed, but few who did not refer to function made this prediction. In contrast, most participants in the accidental context condition thought that the distribution would become more normal if it changed.

To test the relation between historical context, functional explanations, and typicality, we also ran a mediation analysis using the lavaan package in R (Roseel, 2012). For this analysis (not pre-registered), we re-coded the responses for the skewed left distribution so that higher numbers reflected choosing more frequent features for all trials. We then computed a composite score by averaging participants' typicality responses across the four animal trials. Because there were no main nor interactive effects of time in our primary analyses, this analysis collapsed across time periods. The effect of context on typicality judgments was incompletely mediated via functional explanation. As Fig. 7 illustrates, being in the natural context condition was associated with more references to feature function in open-ended explanations ($a = 0.36$; *S.E.* = 0.05; $p < .001$) and referring to function was associated with choosing more skewed category members as typical ($b = 0.22$; *S.E.* = 0.07; $p = .002$); context was associated with typicality indirectly through participants' own explicit function references ($ab = 0.08$; *S.E.* = 0.03; $p = .006$). Importantly, the bias-corrected bootstrapped confidence interval with 10,000 samples was above zero (95% CI [0.03, 0.15]).

We repeated this same procedure for mediation analysis of sample selection responses and found the same pattern. Being in the natural context condition was associated with more functional explanations ($a = 0.36$; *S.E.* = 0.05; $p < .001$) and referring to feature function was associated with choosing more skewed category members to learn from ($b = 0.31$; *S.E.* = 0.10; $p = .001$); context was associated with sample selection indirectly through function references ($ab = 0.11$; *S.E.* = 0.04; $p = .004$). The bias-corrected bootstrapped confidence interval with 10,000 samples was again above zero (95% CI [0.05, 0.20]).

3.2.4. *Do frequency and context shape normative judgments?*

We analyzed participants' responses to all normative claims using linear mixed models (LMMs), with context, time, and claim type included as predictors and including random intercepts for participants and trials; we report the results of Likelihood Ratio Tests.[3] Responses to all normative claims varied by context and judgment, with participants who learned that the animals lived in their natural habitat consistently agreeing with claims that matched frequency—and disagreeing with those that did not match—more strongly than participants who learned that the current distribution was accidental. The context × judgment interaction was significant on all four measures: (a) overall normative *ought* claims, X^2 (1) = 113.66, $p < .001$; (b) teleological norms, X^2 (1) = 99.82, $p < .001$; (c) negative judgments of nonconformity, X^2 (1) = 35.68, $p < .001$; and (d) deontic norms, X^2 (1) = 4.07, $p = .04$ (although agreement with these claims was low overall; see Fig. 8). There were no other significant main nor interactive effects. Individual participants' agreement with all four types of normative claims were highly correlated (all $ps < .001$), however agreement with overall "ought" claims was much more similar to agreement with teleological claims ($M_{diff} = 0.83$) than with negative judgments of nonconformity ($M_{diff} = 1.56$) or deontic claims ($M_{diff} = 1.78$; main effect of claim type, X^2 (2) = 259.96, $p < .001$), confirming that participants in Study 1 likely agreed with *ought* claims because they interpreted them as describing teleological norms.

3.2.5. *Does context shape normative judgments via functional explanation?*

As part of our preregistered analyses, we re-analyzed participants' agreement with normative claims while including their functional explanations as a predictor variable and testing for main and interactive effects. These analyses revealed significant interactive

[3] Our preregistration plan described including distribution as a predictor variable as well; these analyses revealed a variety of interactive effects of distribution reflecting an overall pattern of more extreme responses in skewed right trials than skewed left trials. However, these effects of distribution were orthogonal to our primary hypotheses, and the main patterns of interest were unchanged when distribution was omitted, so for simplicity we report here the results of analyses without distribution included in the models. For a description of the full results as described in our preregistration plan, please see the Supplementary Information on OSF: https://osf.io/863rc/?view_only=28c0c80dfd774c79acd9193043846d38.

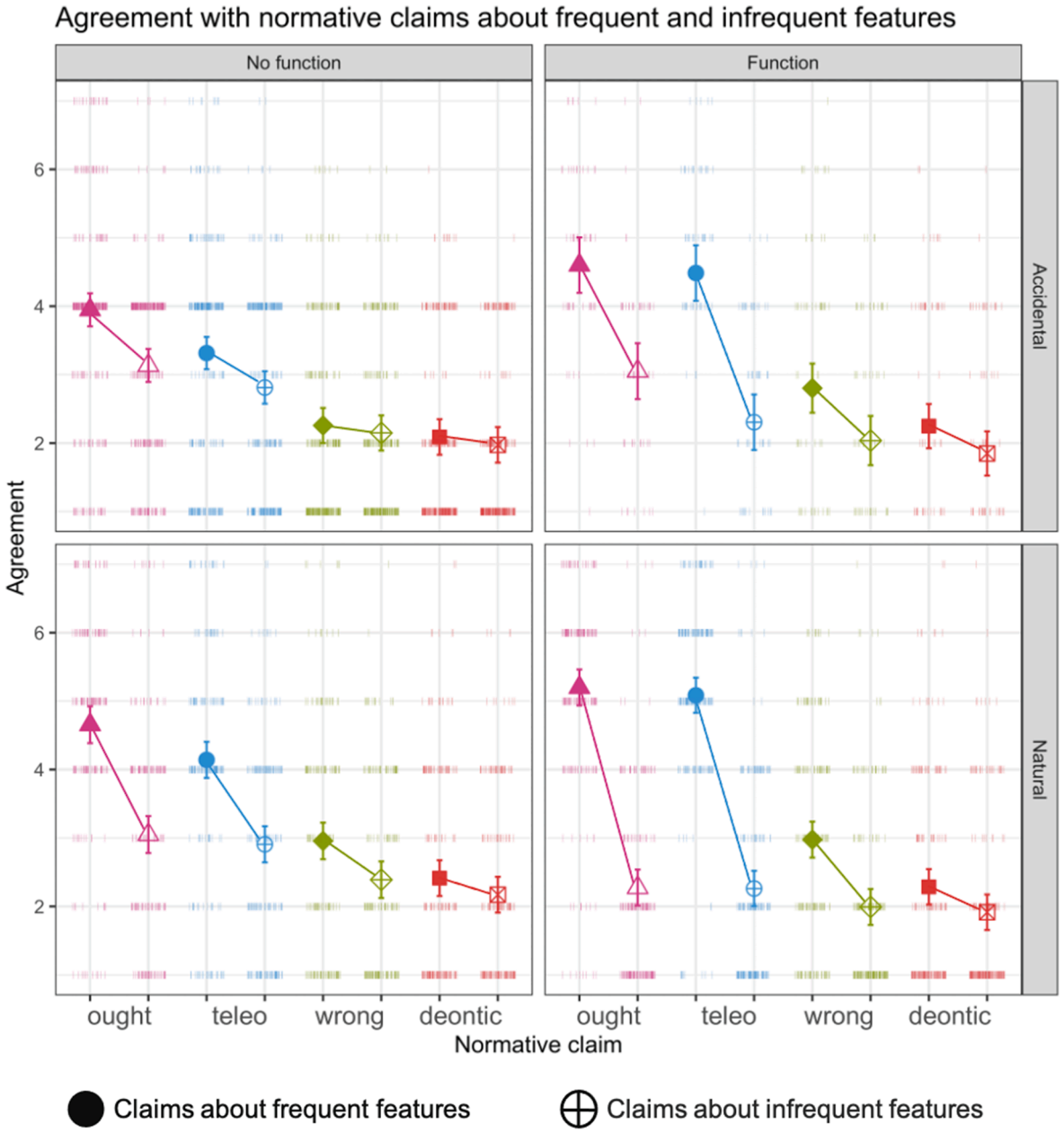

Fig. 8. Study 2: Agreement with four normative claims about frequent and infrequent features (1 = completely disagree, 7 = completely agree), collapsed across trials. Claims included (a) overall *ought* normative judgments, e.g., "Xs ought to have [feature]," (b) teleological norms, e.g., "It would be best for an X to have [feature]," (c) negative judgments of nonconformity, e.g., "There's something wrong with an X without [feature]," and (d) deontic norms, e.g., "An X without [feature] has done something wrong." Filled shapes correspond to claims about frequent features, and hollow shapes correspond to claims about infrequent features, with each pair of claims connected by a line. Panels show the two context conditions (natural and accidental, between participants) and whether participants referred to feature function to explain the distribution (No function = did not refer to function or did not think there was a reason for the distribution, Function = thought there was a reason having to do with feature function). Large shapes show group means with 95% Confidence Intervals, small lines are individual responses.

effects of function reference on all four measures. For (a) overall normative claims, there was a significant two-way function reference by judgment interaction (X^2 (1) = 65.13, $p < .001$), as well as a significant two-way context by judgment interaction (X^2 (1) = 49.86, $p < .001$). Agreement with (b) instrumental/teleological norms showed the same pattern (two-way function reference × judgment interaction, X^2 (1) = 135.89, $p < .001$, two-way context × judgment interaction, X^2 (1) = 28.85, $p < .001$), as well as a significant three-way time × function reference × judgment interaction, X^2 (1) = 5.45, $p = .02$). On both measures, participants who referred to function in their explanations agreed with claims that matched frequency, and disagreed with opposite claims, more strongly than

participants who did not refer to function or who thought there was no reason for the distribution. There was a significant four-way context × time × judgment × function reference interaction for both (c) negative judgments of nonconformity (X^2 (2) = 6.07, p = .002) and (d) deontic norms (X^2 (2) = 4.57, p = .01), as well as several subsumed main and interactive effects.

To test the relation between historical context, functional explanations, and normative *ought* judgments, we ran another mediation analysis using the lavaan package (Roseel, 2012). For this analysis (not pre-registered), we included only agreement ratings for claims about frequent features, and we computed a composite score by averaging participants' agreement across the four animal trials. This analysis was collapsed across time periods. The effect of context on agreement with normative claims about frequent features was again incompletely mediated via functional explanations. Being in the natural context condition was associated with more functional explanations (a = 0.36; *S.E.* = 0.05; p < .001) and referring to feature function was associated with agreeing more with *ought* claims that matched frequency (b = 0.64; *S.E.* = 0.25; p = .01); context was associated with agreement indirectly through functional explanations (ab = 0.23; *S.E.* = 0.09; p = .01). The bias-corrected bootstrapped confidence interval with 10,000 samples was above zero (95 % CI [0.06, 0.43]).

3.2.6. Does context shape normative judgments via typicality?

To test whether functional explanations give rise to normative evaluations *via* representations of what's "normal" for a category (reflected in typicality judgments), we ran a series of exploratory analyses including participants' typicality judgments as a predictor variable. These analyses revealed significant interactive effects of typicality on three out of four normative claim types; for details see the OSI. We also tested whether the effect of context on agreement with normative claims about frequent features was mediated by typicality judgments. We did not find evidence for significant mediation (ab = 0.11; *S.E.* = 0.06; p = .06); however, the bias-corrected bootstrapped confidence interval with 10,000 samples was above zero, 95 % CI [0.03, 0.27]. Together these results suggest that normality may be one avenue through which functional explanations shape normative judgments, but explanations also shape these judgments directly.

3.3. Discussion

In Study 2, the role of feature frequency in shaping participants' category representations depended on participants' explanations for *why* those features became frequent. Participants often believed that features were common among animals living in their natural habitat *because* they served an important function for the animal's survival, and these beliefs mediated both category representations and normative judgments about how animals ought to be. Thus, Study 2 provided clear support for our central hypothesis—that what "is" shapes beliefs about what "ought" to be via people's explanations about *why* things are the way they are. However, we found mixed results regarding the role of people's representations of typicality in normative judgments. These results suggest that the path in our causal model (Fig. 1) from frequency to explanations to normative judgments via category representations represents one of several possible avenues through which descriptive information gives rise to normative judgments. Although we suspect that in everyday cognition people may often base their judgments and predictions on their conceptions of what's normal (which depend on explanations about frequency), the current results show that when functions are made salient people may also explicitly call to mind functional explanations themselves to justify their judgments and predictions, resulting in a direct link from explanations to normative judgments. Given that these analyses were exploratory, future confirmatory research should test the conditions under which functional explanations give rise to normative evaluations directly, versus via category representations and beliefs about what is typical.

We also found clear support for the proposal that functional explanations specifically subserve teleological normative judgments, and that participants in Study 1 likely agreed with overall normative *ought* claims because they interpreted them as describing teleological norms. However, although agreement with claims about deontic norms was low overall, these claims also varied by context, with participants in the natural context condition agreeing with claims that animals had done something wrong if they did not display frequent features, and disagreeing with similar claims about infrequent features, more than those in the accidental context condition. This effect is particularly striking given that claims about deontic norms don't seem applicable to animals who have no control over their physical features, perhaps suggesting that different types of normative judgments are not sharply differentiated in people's commonsense intuitions. We address this possibility further in Study 4.

4. Study 3

The goal of Study 3 was to test the causal role of functional explanations in shaping people's expectations about frequency, their category representations, and their normative judgments. One limitation of Studies 1 and 2 is that the relationships between functional explanations and judgments about typicality and normativity were correlational. In Studies 3 and 4, we provide a stronger test of the hypothesis that functional explanations shape category representations and normative judgments by experimentally manipulating beliefs about feature function. Specifically, in Study 3 we asked whether information about feature function plays a causal role in frequency estimates, typicality judgments, and normative judgments. Including these measures extends previous research showing that people expect functional features to be frequent (e.g., Lombrozo & Rehder, 2012) by asking whether people expect that animals *ought* to display functional features, and by testing how these judgments relate to beliefs about typicality and frequency. Study 3's design and analyses were preregistered on OSF, https://osf.io/ft68m.

4.1. Methods and materials

4.1.1. Participants

Of our recruited 106 adults, 4 were excluded for having an IP address that identified them as being outside of the United States, and we excluded 5 participants for incorrectly answering more than 2 out of 8 attention and manipulation check questions throughout the study, per our preregistration. This left 97 participants (57 female, 39 male, 1 non-binary; $M_{age} = 34$), recruited using Prolific and tested using Qualtrics. Participants were paid $1.50.

4.1.2. Procedure

As in Studies 1 and 2, participants learned about four novel animals, each presented visually with five exemplars that varied along a single feature dimension, such as number of spots. However, participants in Study 3 did not learn anything about the relative

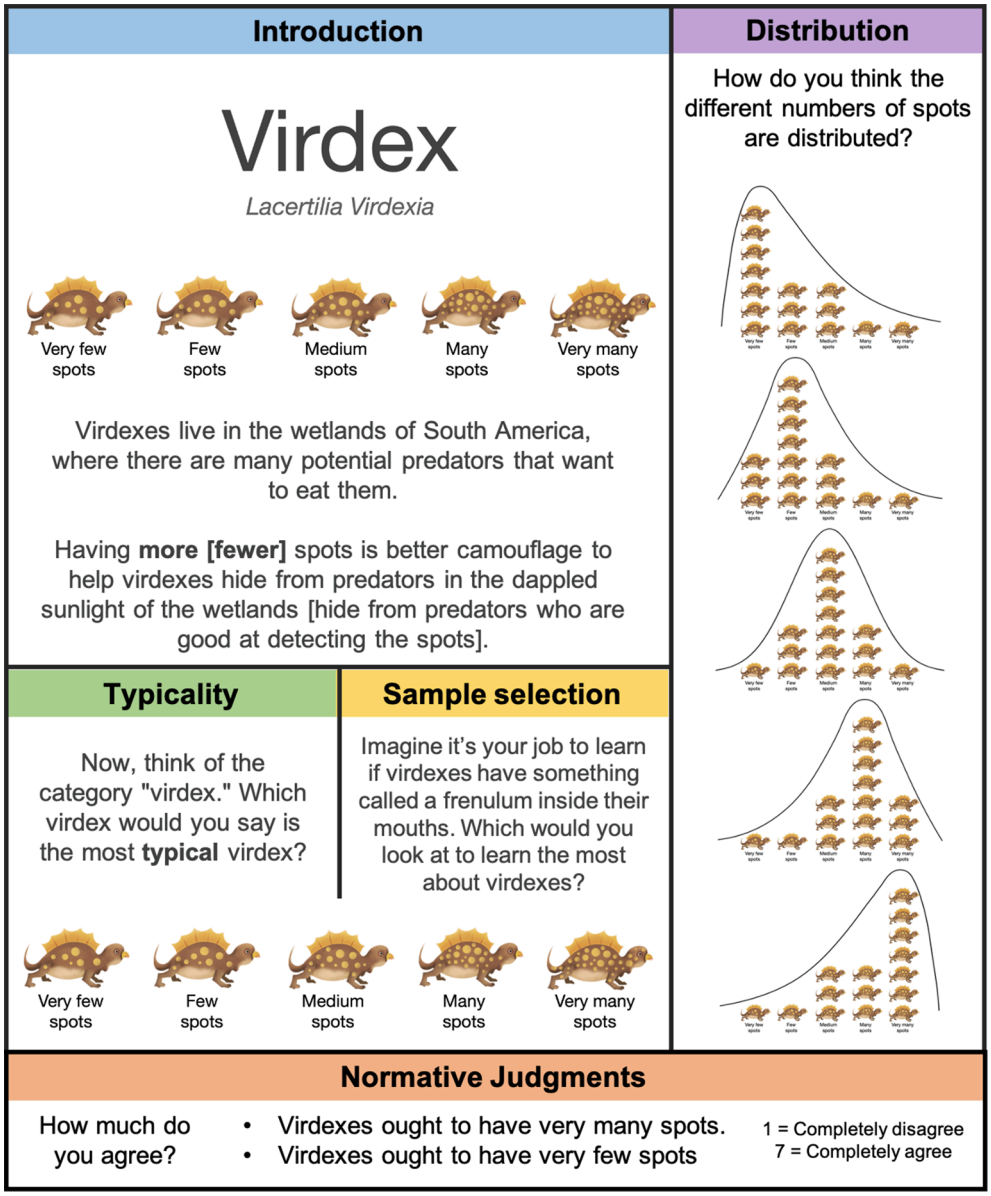

Fig. 9. Method, Study 3. On each animal trial, participants first learned about each animal kind from five category members varying along a single feature dimension. They then answered questions about typicality and informativeness, and questions about feature frequency, in counterbalanced order (between participants). On each measure, participants responded by clicking on the image corresponding to their choice. The entire procedure then repeated for the remaining three animal kinds. After completing all four animal trials, participants evaluated normative claims about all four animals.

frequencies of the different features. Rather, on each trial participants learned that an adaptive fitness function (e.g., better camouflage from predators) was best met by either the left or the right of the scale (e.g., very few spots or very many spots). They then answered a series of questions measuring their beliefs about the category.

To measure participants' expectations about the frequency of functional features, participants picked which feature distribution (of five possible, presented in random left–right or right-left order) best described the category, as in Study 2. Participants also selected which category member they viewed as most representative and informative as in Studies 1 and 2 (Fig. 9). Question order was counterbalanced between participants, with half of participants answering questions about frequency first, and the other half answering questions about typicality and informativeness first.

After completing all four animal trials, we also asked participants how much they agreed with normative claims about how members of each category "ought" to be, as in Study 1. Participants rated how much they agreed with claims about both the most functionally ideal feature values (e.g. "Xs ought to have [most ideal feature]") and about the least ideal features (e.g., "Xs ought to have [least ideal feature]"), presented in counterbalanced order within participants.

4.2. Results

4.2.1. Do functional explanations shape typicality and frequency?

We analyzed participants' typicality, sample selection, and predicted distribution responses using ordinal logistic regression mixed models (CLMMs), with functional ideal condition (left-ideal, right-ideal) as a predictor, and including random intercepts for order, participants, and trials. We report the results of Likelihood Ratio Tests. Participants expected that functionally ideal feature variations would be more common (X^2 (1) = 357, $p < .001$), as well as more typical (X^2 (1) = 361.90, $p < .001$) and informative (X^2 (1) = 393.85, $p < .001$). The interaction between question order and functional ideal condition was significant on all three measures (distribution: X^2 (1) = 24.48, $p < .001$; typicality: X^2 (1) = 10.03, $p = .002$; sample selection: X^2 (1) = 11.25, $p < .001$). Participants who answered distribution questions first gave more extreme responses overall, but the pattern of responses was very similar across question order conditions (Fig. 10). Individual participants' responses across typicality, sample selection, and distribution measures were highly correlated[4] (all $ps < .001$).

4.2.2. Do functional explanations shape normative judgments?

We analyzed participants' agreement with normative claims about how each animal ought to be using linear mixed models (LMMs), including as predictors the functional ideal condition (left-ideal, right-ideal) and judgment type (claims that animals ought to display functionally ideal features, claims that animals ought to display non-ideal features), testing for main and interactive effects. We included random intercepts for order, participants, and trials, and we report the results of Likelihood Ratio Tests. Participants overwhelmingly agreed with normative claims that animals ought to have functionally ideal features (e.g. "Xs ought to have [functional feature]") and disagreed with opposite claims (e.g., "Xs ought to have [opposite feature]"; main effect of judgment type: X^2 (1) = 5301.55, $p < .001$; Fig. 10). In exploratory analyses with typicality judgments included in the model as a predictor, participants who judged more ideal category members as more typical also agreed more strongly with normative claims (three-way functional ideal × judgment × typicality interaction, X^2 (1) = 19.07, $p < .001$). Thus, participants thought that animals *ought* to display functionally ideal features, especially when they viewed those features as representative of what the animals are normally like.

4.3. Discussion

In Study 3, participants expected functional features to be frequent, and they saw category members that displayed them as more typical and informative about their kinds. The fact that responses to all three measures were more skewed when participants answered questions about the expected distribution of features first suggests that reasoning about frequency – specifically, *why* certain features are frequent – might be an important aspect of how frequency shapes category structure. Participants also overwhelmingly thought that animals *ought* to display functional features, as in Studies 1 and 2. This result is in line with our model's prediction that functional explanations license normative beliefs by establishing a standard against which category members can be judged.

Importantly, Study 3 goes beyond Studies 1 and 2 in establishing a causal relationship between functional explanations and judgements about what category members are like, and should be like, complementing the correlational results from Studies 1 and 2. However, as in Study 1, it is unclear precisely how participants in Study 3 were interpreting normative claims about how animals *ought* to be (i.e., as expressing epistemic norms, teleological norms, or deontic norms). Also, like Study 1, Study 3 did not offer direct evidence for the role of folk-biological theories about the causal processes that shape categories over time in moderating effects of functional explanation. We addressed these issues in Study 4.

[4] Our preregistration plans for Studies 3 and 4 described also analyzing participants' typicality and normative judgments with their predicted distribution included as a predictor. For simplicity, we describe these analyses in the Online Supplementary Information on OSF: https://osf.io/863rc/?view_only=28c0c80dfd774c79acd9193043846d38.

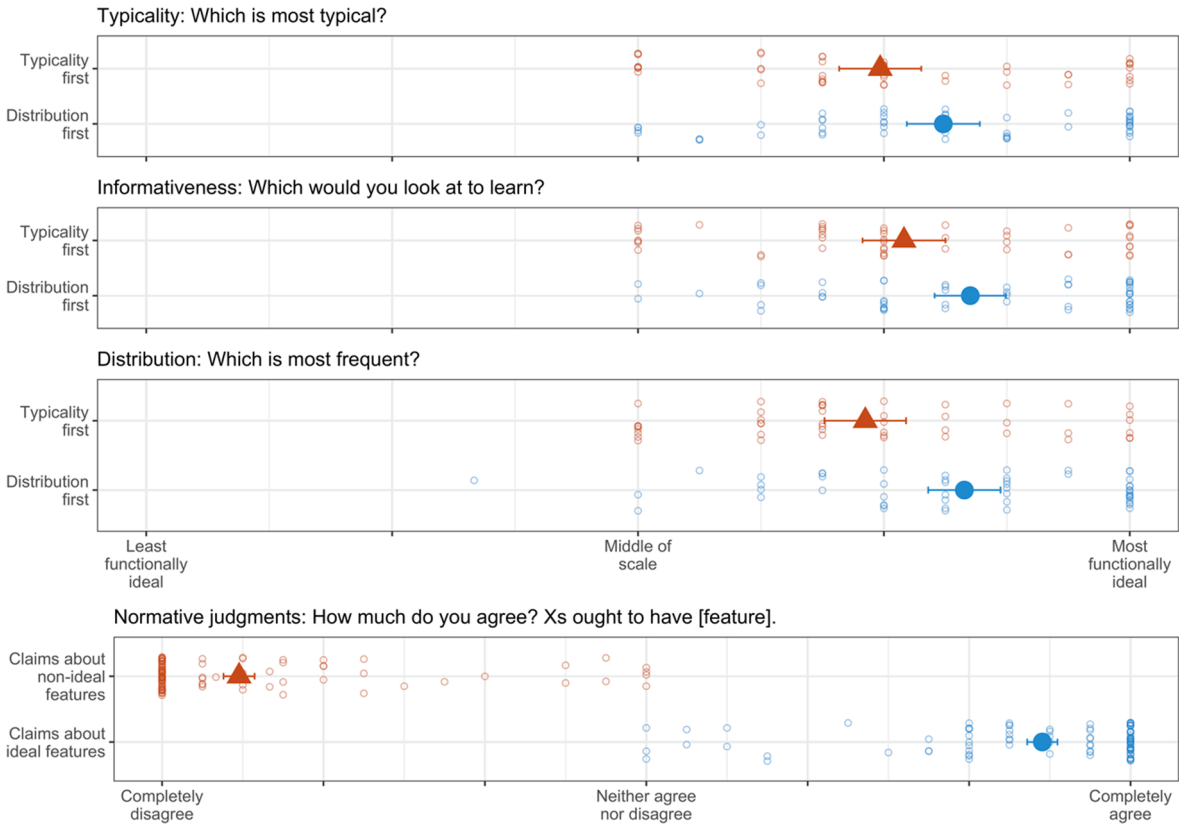

Fig. 10. Study 3: Participants' responses in Study 3. Typicality, sample selection, and expected distribution responses are shown by the order in which participants answered the sets of questions (distribution questions first or typicality and sample selection first, counterbalanced between participants). Responses for the left ideal trials are reverse coded so that the most functionally ideal category member is at the right of the scale for all trials. Agreement with normative *ought* claims are shown by whether the claims describe functionally ideal or non-ideal features. Large shapes show group means with 95% Confidence Intervals, small circles are individual participant averages.

5. Study 4

Our findings so far suggest that functional explanations for biological features predict (Studies 1–2) and shape (Study 3) descriptive expectations about the features' importance for the category, as reflected in judgments of typicality and informativeness. Functional explanations also predict (Study 2) and shape (Study 3) normative evaluations concerning whether category members ought to have those features, most likely by establishing a teleological norm (Study 2). Finally, these effects reflect people's intuitive theories of biological change: high frequency more often prompts functional explanations when the frequency resulted from natural biological processes, and in these natural contexts, functional explanations predict expectations of future change (Study 2).

Study 4 goes beyond these results by testing the following implication. If effects of functional features depend on people's intuitive theories of biological change in "natural" conditions, as reflected in their functional explanations, then it should be possible to break or attenuate the link between functional features, typicality, and normative judgments by presenting functional features that do not support functional explanations. For example, although holding up glasses and supporting respiration are both functions of the nose, only its role in respiration was instrumental in perpetuating noses: We have noses because they help us breathe, not because they hold up our glasses**.** Both normative and descriptive accounts of functional explanation suggest that functional explanations are more appropriate when the cited function is "historical" – that is, when it played a causal role in bringing about or perpetuating the feature, like respiration for noses – versus "ahistorical" – that is, when it is merely a consequence of some feature or system, like holding up glasses (Lombrozo & Carey, 2006; Wright, 1976). Thus, it should be possible to break or attenuate effects of functional features by presenting features with ahistorical functions (such as holding up glasses), as opposed to historical functions (such as supporting respiration).

In Study 4, all participants learned about functions that features currently serve (e.g., that many spots are good for camouflage), but in the natural context these functions were historical, and in the accidental context these functions were ahistorical. Although both functions establish a normative standard, we expected that historical functions would have larger effects on typicality and normative evaluations. For instance, people might agree with normative claims about historical functions (e.g., "noses ought to help us breathe") more than claims about ahistorical functions (e.g., "noses ought to hold up our glasses"). This finding would suggest an important

boundary condition on the effects of functional features, as well as providing additional evidence for the moderating role of intuitive theories and causal commitments in driving inferences from *is* to *ought* (see Fig. 1).

We also further examine the possibility that functional reasoning (even about ahistorical functions) can guide reasoning about how animals might change in the future. Although typicality judgments did not vary for the present and future in Study 2, one interesting result of Study 2 was that participants in the natural context condition—and especially those who referred to feature function to explain current frequency—were more likely to predict that populations of animals would become more skewed towards functional features in the future, rather than less skewed (i.e., more normal). This result is consistent with the causal model described in section 1.1: As long as the causal processes that led functional features to become frequent are still at play, then people will expect those processes to cause functional features to continue changing in the same direction as time progresses. This pattern is also in line with previous evidence of a teleological bias, or the persistent tendency to misunderstand evolution as getting better over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014).

Study 4 provides a direct test of how functional information affects expectations about future versus present prevalence, as participants were provided with information about feature function but no information about frequency, and we compared judgments about expected frequency in the present versus the future. We designed Study 4 to have the same structure as Study 2, with the intention of comparing across studies. Study 4′s design and analyses were preregistered on OSF, https://osf.io/5g2x6.

5.1. Methods and materials

5.1.1. Participants

Of our recruited 212 adults, 8 were excluded for having an IP address that identified them as being outside of the United States, and 4 were excluded for incorrectly answering more than 3 out of 14 attention and manipulation check questions throughout the study. This left 200 participants (115 female, 80 male, 3 non-binary, 2 who preferred to self-identify; $M_{age} = 37$ years), recruited using Prolific and tested using Qualtrics. Participants were paid $1.90.

5.1.2. Procedure

The design of Study 4 mirrored that of Study 2, using a 2 (context: natural, accidental) × 2 (time period: current, future) × 2 (functional ideal: left-ideal, right-ideal) design, with all conditions assigned randomly by Qualtrics. As in Study 2, historical context

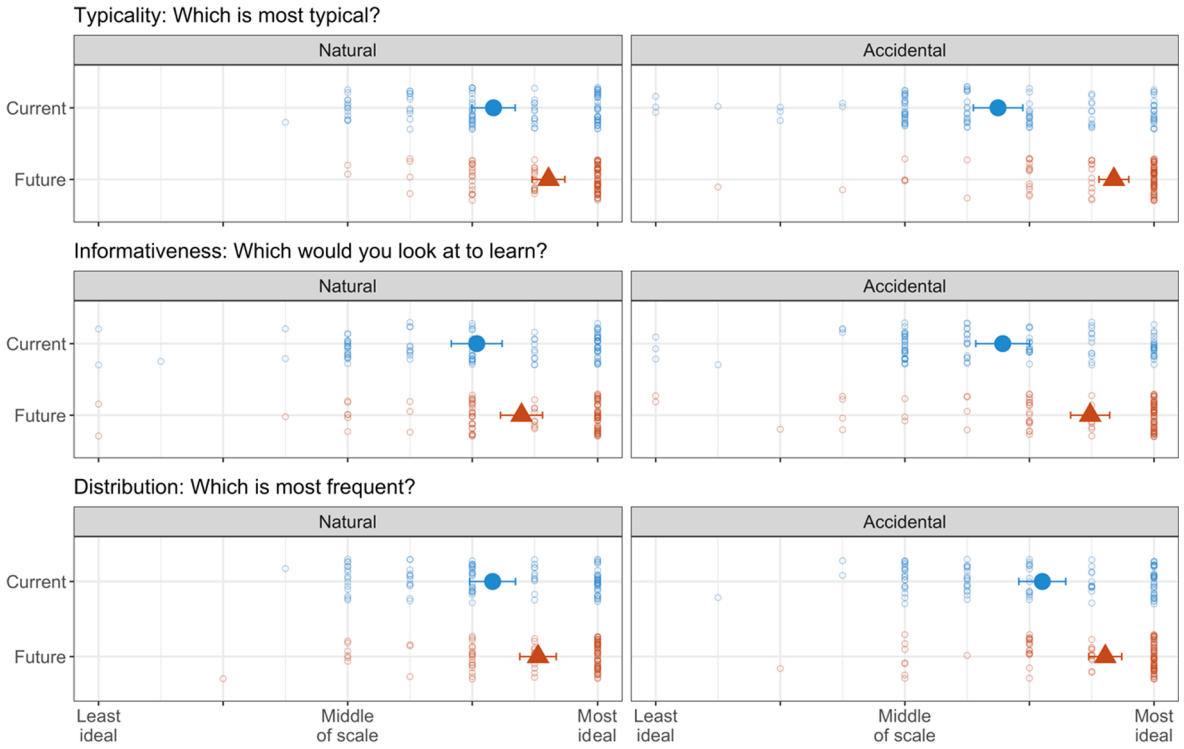

Fig. 11. Study 4: Participants' typicality, sample selection, and distribution responses, collapsed across trials. Responses for the left ideal trials are reverse-coded so that the most functionally ideal category member is at the right of the scale for all trials. Panels show the two context conditions (natural and accidental) and time periods (current and future, both within-participants). Large shapes show group means with 95% Confidence Intervals, small circles are individual participant averages.

was between-participants: Half of participants learned that all four animals lived in their natural habitat undisturbed by humans (natural context), and that certain features were more functionally ideal within this habitat. The other half learned that, due to accidental factors, all four animals had recently been introduced into environments that were very different from their natural habitat (accidental context) and that certain features just happened to be functional in the new environment. These features' functions were thus ahistorical in the present because they could not have played a causal role in bringing about or perpetuating the features, but they had the potential to *become* historical in the future (this would be the equivalent of noses' role in holding up glasses coming to exert

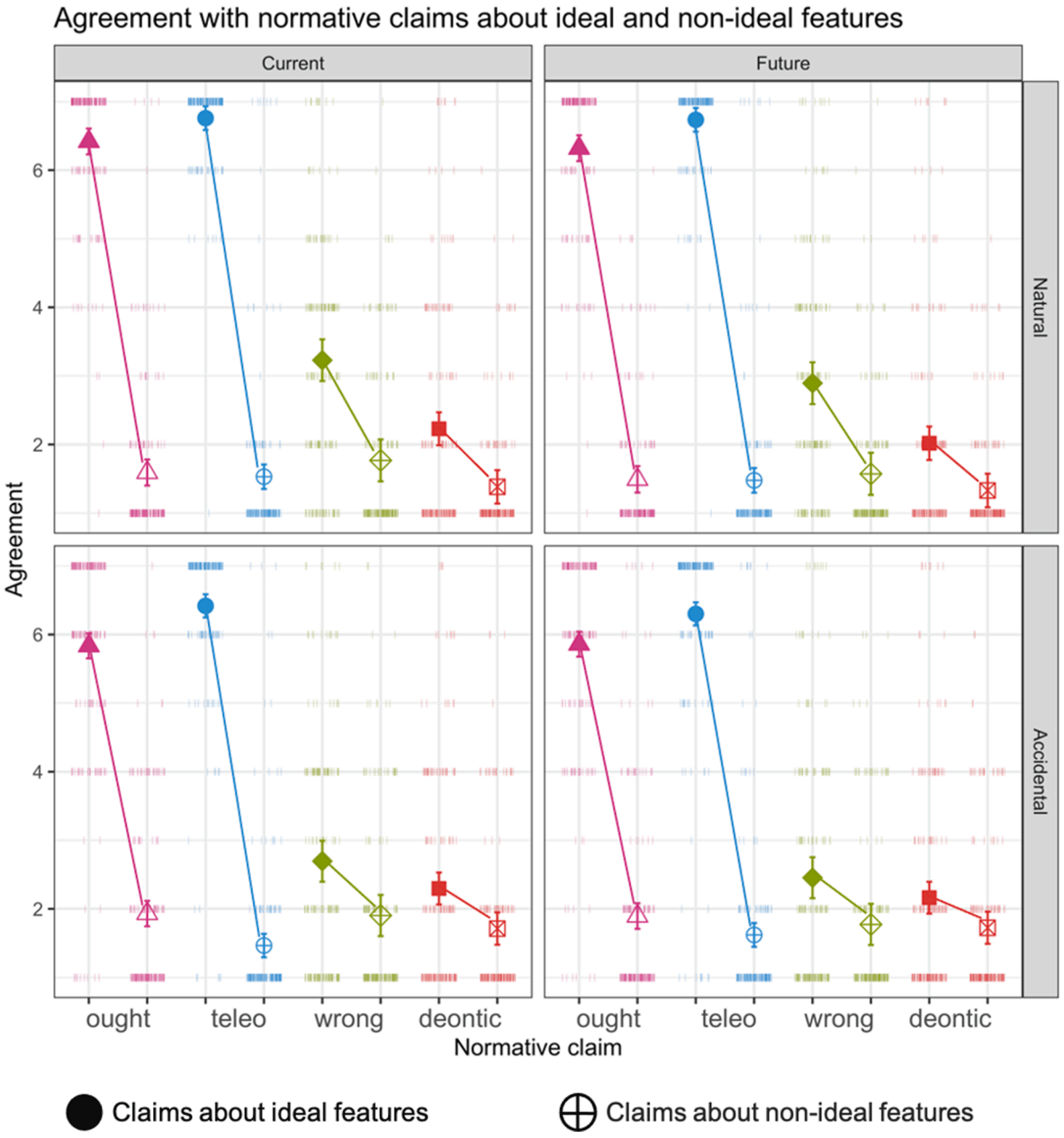

Fig. 12. Study 4: Agreement with four normative claims about functionally ideal and non-ideal features (1 = completely disagree, 7 = completely agree), collapsed across trials. Claims included (a) overall *ought* normative judgments, e.g., "Xs ought to have [feature]," (b) teleological norms, e.g., "It would be best for an X to have [feature]," (c) negative judgments of nonconformity, e.g., "There's something wrong with an X without [feature]," and (d) deontic norms, e.g., "An X without [feature] has done something wrong." Filled shapes correspond to claims about ideal features, and hollow shapes correspond to claims about non-ideal features, with each pair of claims connected by a line. Panels show the two context conditions (natural and accidental) and time periods (current and future, both between participants). Large shapes show group means with 95% Confidence Intervals, small lines are individual participant averages. All ideal vs non-ideal contrasts were significant and all means differed significantly from the mid-point.

selective pressure on the maintenance or properties of noses).

Also as in Study 2, time-period was within-participants: For the first two trials, all participants answered questions about the animals as they currently are, and for the third and fourth trials they made judgments about animals 100 generations in the future. Functional ideal condition was counterbalanced within-participants using a Latin square design: On each of 4 animal trials, participants learned that the most functionally ideal feature value was either on the left of the scale (left-ideal, 2 animal trials) or the right of the scale (right-ideal, 2 animal trials; modified so that all participants saw one of each functionally ideal condition in each time-period block). All participants answered typicality and sample selection questions first, followed by predicted distribution questions (recall that in Study 3 this question order led to less extreme responses).

5.2. Results

5.2.1. Do function and context shape typicality and frequency?

Typicality responses showed a significant three-way context × time × functional ideal interaction (X^2 (1) = 12.93, $p < .001$), as well as several subsumed main and interactive effects. For animals in the present, participants in the natural context chose more functionally ideal category members as typical than those in the accidental context; those in the accidental context chose category members closer to the middle of the scale as typical (i.e., 3; pairwise contrasts: left-ideal, $p = .002$; right-ideal, $p = .001$). However, participants in both conditions thought that functionally ideal category members would be similarly typical 100 generations in the future (left-ideal, $p = .74$; right-ideal, $p = .26$), once the ahistorical functions had time to *become* historical. The same pattern was also found for sample selection responses (three-way context × time × functional ideal interaction: (X^2 (1) = 5.82, $p = .02$). Distribution responses, in contrast, showed only a significant two-way time × functional ideal interaction (X^2 (1) = 54.02, $p < .001$), with participants predicting that functionally ideal category members would be frequent in the present—and become more frequent in the future—to a similar extent across both contexts. The finding that context affected typicality judgments, but not beliefs about frequency, illustrates how representations encode not only information about which features are frequent, but also filter this information through beliefs about *why* they are frequent. Nevertheless, individual participants' responses across measures were highly correlated (all *p*s < .001; Fig. 11).

5.2.2. Do function and context shape normative judgments?

As in Study 2, responses to all normative claims varied by context and judgment, with participants who learned that the animals lived in their natural habitats consistently agreeing with claims about functionally ideal features, and disagreeing with claims about non-ideal features, more strongly than participants who learned that animals lived in a new environment (Fig. 12). The context × judgment interaction was significant on all four measures, but showed the strongest effects on (a) overall normative claims, X^2 (1) = 47.61, $p < .001$, and (b) instrumental/teleological norms, X^2 (1) = 13.65, $p < .001$ (for further details, see the Online Supplementary Information).

5.2.3. Does typicality predict normative judgments?

To test whether typicality predicts normative judgments of individual category members, we again conducted a series of exploratory analyses of participants' normative judgments (about ideal features only) with typicality responses included as a predictor, along with context and time, testing for main and interactive effects. Participants in the natural context who viewed more ideal category members as typical also agreed more strongly with (a) overall normative claims (context × typicality interaction, X^2 (1) = 15.87, $p <$ .001) and (b) instrumental/teleological norms (context × typicality interaction, X^2 (1) = 6.03, $p = .01$; subsumed main effect of typicality, X^2 (1) = 6.95, $p = .009$) whereas typicality did not predict normative judgments for participants in the accidental context (for further details see the OSI). These results support the possibility that one route through which causal beliefs and feature function shape normative judgments is via representations of what categories are normally like.

5.2.4. Comparison of Studies 2 and 4

As described in our Study 4 preregistration document, we planned to compare results across Studies 2 and 4 to test whether functional information plays a particularly central role in licensing predictions about future change. If frequency (manipulated in Study 2) influences typicality and normative claims through functional explanations (manipulated in Study 4), we would also expect more extreme responses in Study 4 than Study 2, at least in the natural contexts. This is precisely what we found: compared to frequency information (Study 2), functional information (Study 4) had larger and more reliable effects for the future compared to the present and resulted in more extreme responses on all measures in natural contexts (see OSI for analyses.).

5.3. Discussion

The results of Study 4 highlight an important boundary condition on the effects of functional features and provide additional evidence for the role of intuitive theories and causal commitments in driving inferences from *is* to *ought*. First, replicating Study 3, Study 4 found that functional features were judged more prevalent, and that they supported normative claims that category members *ought* to possess functional features. Consistent with the findings from Study 2, these normative claims corresponded most closely to endorsement of teleological norms. Second, going beyond Study 3, Study 4 found that these inferences were moderated by the causal status of a function. When a function was historical (the equivalent of the nose's role in respiration) versus ahistorical (the nose's role

in holding up glasses), participants viewed more skewed features as typical, and they more strongly endorsed normative claims. However, when asked to consider feature distributions 100 generations in the future, even ahistorical functions supported strongly skewed predictions, suggesting that participants expected ahistorical functions to effectively *become* historical.

The comparison between Studies 2 and 4 further suggests that functional information – more so than current frequency – shapes expectations about how species will change in the future. Functional explanations reflect commitments regarding the causal processes that shaped category members over time, so people will expect functional features to continue becoming more frequent in the future as long as those same causal processes are still at play. This pattern is in line with previous evidence that people tend to misunderstand evolution as getting better over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014). At the same time, the contrast between historical and ahistorical functions in the present helps make sense of prior work suggesting that people view patterns that have existed for longer as holding more normative force (i.e., the Longevity Bias; Blanchar & Eidelman, 2021; Eidelman & Crandall, 2012; Eidelman et al., 2010): this could be because they are more likely to be regarded as reflecting historical functions.

6. General discussion

6.1. Folk theories underlie is-ought reasoning

Across four studies, we found evidence in support of the proposal that a key mechanism through which *is* shapes *ought* for folk-biological concepts rests on people's causal explanations about *why* certain features became frequent to begin with. The current results provide support for each of the theoretical predictions in our causal model (Fig. 1): Participants tended to spontaneously explain patterns in nature by appealing to feature function (e.g., reasoning that virdexes have many spots because spots serve an important function, like camouflage; Studies 1 and 2), and they likewise inferred that functional features would be frequent (i.e., if having many spots is functional, then most virdexes will have them; Studies 3 and 4). These explanations partially predicted participants' representations of what is typical, or normal, for the categories (Studies 1 and 2), and in fact played a causal role in shaping these category representations (Studies 3 and 4). By setting a normative standard against which category members could be judged as better or worse (e.g., in terms of potential for camouflage), functional explanations also licensed normative judgments about what category members *ought* to be like (Studies 1–4). Finally, participants expected that functional features would become even more frequent in the future (Studies 2 and 4), in line with prior evidence that people's folk-biological expectations about nature entail thinking that evolution means getting better over time (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014).

Our proposal is consistent with broader theories that concepts are structured by people's causal-explanatory frameworks (Ahn, 1998; Ahn et al., 2000; Carey, 1985; 2000; Gopnik & Meltzoff, 1997; Lombrozo & Carey; 2006; Murphy & Allopenna, 1994; Murphy & Medin, 1985; Rehder & Kim, 2010; Rips, 2011; Waxman & Gelman, 2009; Wellman & Gelman, 1998), as well as with previous suggestions that is-ought reasoning arises from people's explanations about how features relate to their kinds (Haward et al., 2018; Prasada & Dillingham, 2006; 2009; Tworek & Cimpian, 2016). However, to our knowledge, the current proposal is the first to combine these bodies of work, providing evidence that, at least within the biological domain, is-ought reasoning arises specifically from people's causal explanations about the processes that led features to become common over time.

Nevertheless, our proposal is not inconsistent with previous suggestions about the importance of explanations for is-ought judgments, and it may well be that several different forms of explanation license corresponding normative claims. For example, Tworek and Cimpian (2016) propose that people tend to rely on explanations that appeal to inherent causes, and these inherent explanations bring with them a sense of necessity that in turn licenses normative claims. However, it is not entirely clear why inherence or necessity would have this consequence, nor what form of normativity would follow. In light of the current results, one possibility is that the inherent explanations people posit may tend to appeal to function, either directly or by citing the proximate causes of functional properties; future research should explore this possibility.

Prasada and colleagues (Prasada & Dillingham, 2006; 2009; see also Haward et al., 2018; 2021) suggest that category-based explanations for a given feature (e.g., that a given animal barks *because* it is a dog) offer evidence that the category and feature share a "principled connection," meaning that the feature is seen as an aspect or *part* of being a member of that kind. This part-whole relationship plausibly licenses the normative judgments they document (e.g., that there is something wrong with a dog that doesn't bark, just as there is something wrong with an object missing a part). However, beyond a general reliance on prior beliefs, it is unclear how people learn which properties are principally connected to their kinds (Haward et al., 2021). One possibility is that people tend to view features that support functional explanations as sharing a principled connection to their kind (Prasada & Dillingham, 2006; 2009). For example, people tend to accept teleological generic statements that describe principled connections (e.g., "cars are for driving") but reject teleological generic statements about non-principled connections (e.g., "cars are for parking"; Korman & Khemlani, 2020). In line with this possibility, participants in Study 4 agreed more strongly with normative claims about historical functions (which supported functional explanation) than with normative claims about ahistorical functions (which just happened to be beneficial for animals in a new habitat).

Finally, Lewry aet al. (2021) found that functional explanations for an entire *species* (e.g., that humans exist to reproduce) are taken to imply that the function is good for the species as a whole, with the consequence that individuals who choose not to fulfill their species function are immoral. Given that a teleological norm does not necessarily license a moral claim (e.g., "zebras who don't have stripes are immoral"), it seems likely that the findings from Lewry et al. reflect features beyond those in our own studies. For instance, many participants in Lewry et al. assumed that the species' functions emerged from individual or group choice (not just natural

selection), and the targets of normative evaluation were intentional agents who chose not to act in accordance with their species function, not category members who had or lacked physical features.

That said, participants in our studies were more likely to endorse claims about deontic norms when they viewed frequent features as functional within the context of natural causal processes – either because they had generated functional explanations themselves (Study 2), or because we had explicitly manipulated which functional explanations were warranted (Study 4). One interpretation of this finding is that different types of normative judgments are not as cleanly differentiated as one might expect. That is, participants' deontic judgments about wrongdoing may be subtly influenced by other normative judgements, for example because of negative affective responses to the violation of a category goal. In line with this interpretation, some researchers have proposed that people hold undifferentiated views of what is probable and what is morally right (Phillips & Cushman, 2017). However, the observed effects of context on deontic judgments were small, and it could be that participants were simply motivated to give consistent ratings across the four types of normative statements. An important direction for future research, then, is to expand the approach we take in Studies 2 and 4, probing a range of normative judgments (teleological, deontic, etc.) to examine how they relate, and how they depend on people's beleifs about the causal processes that gave rise to the function in question.

6.2. The "natural" and the normative

In the context of biological kinds, views of what is typical or normal may be more accurately described as views of what is "natural," because people expect the rules that govern natural kinds to be set by nature (rather than humans). Prior work has already shown that people tend to believe that "natural" means "good," suggesting that the phenomena we document here might fall under the broader heading of a *naturali*stic fallacy that takes the natural seriously. For example, consumers are often willing to pay more for foods labelled "natural" and many oppose genetically-modified foods because they are judged to be "unnatural" (Rutjens et al., 2018; Scott & Rozin, 2020). In the social domain, people often appeal to nature to justify their normative judgments about same-sex marriage (O'Connor, 2017; Rozin, 2005) and violations of gender roles (Brescoll & LaFrance, 2004). Indeed, Daston (2014) describes the naturalistic fallacy itself as "a kind of covert smuggling operation in which cultural values are transferred to nature and nature's authority is then called upon to buttress those very same values" (Daston, 2014, p. 580).

But why would the "natural" have any normative implications? Our proposal suggests that the natural will license the normative when the natural is explained teleologically. This could arise from belief in divine creation but is also consistent with common beliefs about nature. For instance, many people think of nature as agentive, so they view natural progress over time as akin to the goal-directed actions of an agent (i.e., Gaia beliefs; Blancke et al., 2014; Järnefelt et al., 2015; Kelemen, 2012, Moore et al., 2002). Thus, people might explain the features of natural kinds in terms of functions, and think that they *should* fulfill those functions, because it is the intention of their creator (whether God or nature itself) that they should do so. But teleological beliefs about evolution may be sufficient to support a pervasive tendency to explain the biological world by appeal to function, even without an explicit appeal to God or Gaia (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014). While people tend to think of evolution as being more teleological than it is, many biologists and philosophers of biology in fact accept teleological claims regarding the products of natural selection (for relevant discussion, see Lombrozo & Carey, 2006).

One implication of the current proposal is thus that people's is-ought reasoning could be thought of as the product of an ultimately rational process, but one that happens to be operating with assumptions about the world that are often mistaken (e.g., thinking of nature as a goal-oriented agent, or thinking of all biological features as adaptations). That is, although the conclusions that people arrive at through these reasoning processes may sometimes be mistaken, they engage in rational reasoning about causal forces, and the mistake arises from inaccurate inputs to the rational system (e.g., that a given feature is an adaptation, when in fact it is a spandrel or the product of culture, or that biological and cultural evolution necessarily entail improvement over time). For example, people might reason that mothers *ought* to be primary caregivers instead of fathers (a mistaken conclusion) because the current gender division is the result of "natural" processes that have shaped human social structures over time (an accurate causal reasoning process), and natural processes entail improvement over time (an inaccurate representation of nature). A different error that people could be making is in failing to differentiate normative entailments: While it may be appropriate under some conditions to endorse a teleological norm (e.g., that, from the perspective of a Zebra, stripes are indeed best), natural selection does not support deontic norms (e.g., that a Zebra without stripes has *done* something wrong).

In Study 1, we attempted to measure variation in folk biological beliefs, with the hope that we might identify beliefs about biological change that moderate effects of feature frequency on normative judgments. However, the folk biological beliefs that we probed concerned participants' basic understanding of evolution; we found little variability and no influence on factors of interest. Future research should focus on more fine-grained aspects of biological change. One candidate is people's beliefs about nature as an agentive force (i.e., their endorsement of teleological explanations for natural processes; Kelemen, 1999; Roberts et al., 2021), or their tendency towards adaptationist thinking. Another possible dimension of variation is the extent to which people believe natural selection occurs relatively unfettered, versus in a highly historically contingent manner. For instance, the fact that stripes are good camouflage for zebras is highly contingent on extrinsic factors, including the visual characteristics of the zebra's environment and the sensory capacities of its primary predators. Under other conditions, other defensive strategies would be more ideal. How participants evaluate the ideal, and what is "close" versus "far" from ideal, is therefore likely to depend on how participants represent such historical contingencies and their alternatives (for relevant discussion of such distance metrics that depend on norms, see Kahneman & Miller, 1986; also Lagnado et al., 2013, Knobe, 2010; Phillips et al., 2015).

6.3. The normative dimensions of conceptual representation

Our findings contribute to a growing body of research showing that people's mental representations of what is typical or normal are not only shaped by descriptive information about statistical frequency, but also by causal-explanatory beliefs (Murphy & Allopenna, 1994; Murphy & Medin, 1985) and normative ideals (Barsalou, 1985; Bear & Knobe, 2017; Bear et al., 2020; Foster-Hanson & Rhodes, 2019a; Phillips & Cushman, 2017; Wysocki, 2020). In the current experiments, category members with features that better instantiated the ideals established by functional explanations were judged as more typical and more informative about their kinds. This result replicates past work in showing a link between functional features and typicality (Foster-Hanson & Rhodes, 2019a), but goes beyond it by showing that functional explanations themselves shape typicality, above and beyond downstream inferences that functional features are likely to be frequent (Lombrozo & Rehder, 2012).

Because typicality is shaped by causal-explanatory beliefs and ideals, we should expect typicality to depart from objective frequency. This is what we see in Studies 1, 2, and 4, and it is also a salient feature of "normality." For instance, it is "normal" for tadpoles to grow into frogs, even though most tadpoles get eaten before they are able to do so (McGrath, 2005, p. 139). The finding that typicality does not only encode something like a central tendency or average has important consequences for people's judgments, because typical or "normal" category members serve as the basis of comparison for other members of the category (Osherson et al., 1990). In this way, people's beliefs about what is typical or normal for a category provide the starting point from which deviations can be judged (McGrath, 2005; Wysocki, 2020).

But why might people's concepts combine both descriptive and normative information in this way? Our findings hint at one possibility: participants predicted not only that functional features were more frequent in the present, but also that they would *become* more frequent over time (Studies 2 and 4; see also Lombrozo & Rehder, 2012). It could be that descriptive information captures the present, while normative information indicates the direction of change, and thus supports inferences about the future. Previous accounts have suggested that normative information about *social* categories may be important for predicting the future behavior of social agents (Del Pinal& Reuter, 2017), and indeed normative ideals serve an important social function, by motivating and constraining human social behavior in groups (Tomasello, 2020; see also Shaw & Blakey, 2020). However, we are unaware of any previous proposals that normative information may be important for prediction in the context of traits and behaviors that are outside of the control of human agents. One possibility is that what unites both folk biological kinds and social kinds is an intuitive theory that category change is goal-directed—in the case of social kinds because change is driven by intentional agents; in the case of biological kinds because change is believed to result from a divine creator, a Gaia-like version of nature, or a goal-directed mischaracterization of evolution (Coley & Tanner, 2015; Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo et al., 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014).

A related but different proposal about why typicality structure might encode both descriptive and normative dimensions comes from their roles in *prediction* versus *action*. When we have to predict the characteristics or behaviors of a category member, we should arguably rely on descriptive information. But when we have to intervene—to modify an object, select an individual, or punish an agent—it makes sense to consult normative considerations as well. Typicality could combine the most likely with the most ideal as a compromise or "all-purpose" representation (see Bear et al., 2020, for related discussion).

How do functional explanations fit in this merger between the descriptive and the normative in category representations? Our causal model (Fig. 1) suggests that functional explanations shape category representations (thus defining the typical or "normal"), and that these representations can in turn support normative judgments. Our studies offer some support for this proposal: Participants' typicality judgments predicted their agreement with normative claims across all four studies, and in Study 2, the effects of functional explanation on normative claims were partially mediated by judgments of typicality. On the other hand, even when typicality was included in our statistical models, functional explanations remained a significant predictor of participants' normative judgments in Studies 2–4. Generating explicit causal explanations is cognitively demanding, so one interpretation of these results is that people might tend to rely on their existing representations of what's "normal" to make everyday judgments, but that functional explanations can also shape people's judgments directly, especially when made salient by context (e.g., when people are given functional explanations or generate their own, as in the current studies). Future research will be needed to identify the conditions under which functional explanations license normative judgments *via* typicality, directly, or both (our own bet is "both").

The current proposal is also consistent with previous evidence that even young children's concepts rely on beliefs about function (Foster-Hanson & Rhodes, 2019a), while suggesting a new interpretation of children's focus on characteristic or principled features (Foster-Hanson & Rhodes, 2022; Haward et al., 2018). That is, young children might be particularly likely to implicitly assume that frequent features are functional, or "on purpose," (Kelemen, 1999), but they might also rely less on explicit causal explanations about exactly *how* they are functional compared with adults because such explanations are costly to generate. Similarly, adults learn to balance their intuitive beliefs in purpose against other learning strategies, so they might be less vigilant about unwarranted teleological explanations when under cognitive load (Kelemen & Rosset, 2009; Roberts et al., 2021). Future research should directly test these predictions about the cognitive and developmental origins of is-ought reasoning.

6.4. From biological to social domains

Views about the natural and the "normal" can have pernicious consequences when applied to categories of people – both in terms of excusing "natural" negative behaviors (e.g., male infidelity or even rape; Dar-Nimrod et al., 2011; Ismail et al., 2012) and in terms of weakening agency and praise for positive behaviors (e.g., mothers having children at immense personal cost) or prohibiting agents from engaging in actions that might go against the "natural order" (e.g., abortion; Wunderlich, 2020). Future work should directly

examine the causal role of beliefs about what is "natural" as a potential mechanism through which people's knowledge of what is common can give rise to normative judgments about both blame and praise in the social world. Given these consequences for social judgments, future work should also test how flexible people are in their reliance on beliefs about nature when making causal attributions. That is, does explaining actions by appealing to the broader causal processes that gave rise to them (e.g., nature, evolution), versus the goals or desires of the agent, constitute a particular stance that people can adopt flexibly or a fixed understanding of the world (Nagel, 1979; Strawson, 1962)? If the former, then interventions that change people's beliefs in the benevolence of nature could be a fruitful tool for social change.

Understanding what shapes people's representations of what is normal is also important because idealized or biased representations can perpetuate cultural ideologies and shape social cognition in subtle ways. For example, despite the fact that dark skin and brown eyes have been the most frequent phenotypes across human history, light skin and eye color (features historically common only among Northern Europeans but viewed as more "ideal" in pro-White western society) are often represented as "normal" in medical textbooks (Graves, 2021). These biased representations shape how people think and reason about the social world beginning in childhood. For example, children begin to view White men, boys, and girls as more "normal" for their social categories beginning as early as third grade (Lei et al., 2021). For these reasons, children's media featuring all or mostly White characters may inadvertently perpetuate pro-White bias (Roberts & Rizzo, 2021). Indeed, the phrase "representation matters" has become prevalent across all facets of mainstream media (Lemish & Johnson, 2019). Yet it remains unclear whether equalizing representation alone – in the absence of explanatory information about *why* the representation was unequal to begin with – would have any measurable effect on children's (and even adults') beliefs about the social world (see also Vasilyeva et al., 2018; Vasilyeva & Lombrozo, 2020). Future work should directly test the causal role of representations, with and without causal explanations, on people's judgments and attitudes.

6.5. Conclusions

The current work proposed a novel mechanism through which descriptive information about how the world *is* can shape people's beliefs about how it *ought* to be. When people think of a typical zebra, what comes to mind is shaped by what is frequent, as well as people's beliefs about what is ideal relative to the species' functional goals; these idealized representations give rise to the belief that zebras *should* have stripes. This work thus sheds light on the two puzzles raised in the introduction. First, where do normative judgments about how the natural world *ought* to be come from? We suggest that these judgments come from people's folk-biological beliefs about the causal forces that shape nature over time, and the expectation that natural progress means improvement. These are reflected in functional explanations that in turn shape judgments of typicality and normativity. Second, how and why do normative judgments relate to our representations of how the world *is*? We propose that these judgments arise from our very representations of what is "normal," which become idealized when descriptive information is filtered through our causal-explanatory frameworks about *why* certain features became frequent over time. Thus, the missing premise from *is* to *ought* arises from people's folk biological beliefs reflected in – and perpetuated by – functional explanations, which are encoded in our representations of the world.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data and analysis code are available on OSF and linked in the ms. (https://osf.io/863rc/)

Acknowledgments

We thank our colleagues in the Princeton University Center for Human Values for their helpful feedback on a previous version of this paper, and we are grateful to the University Center for Human Values and the Program in Cognitive Science for supporting EFH's postdoctoral position. We also thank Casey Lewry for assistance with data coding.

This work was supported by a grant from the National Science Foundation [SMA-1948630].

References

Ahn, W. K. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition, 69*(2), 135–178. https://doi.org/10.1016/S0010-0277(98)00063-8

Ahn, W. K., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*(4), 361–416. https://doi.org/10.1006/cogp.2000.0741

Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language, 55*(3), 402–421. https://doi.org/10.1016/j.jml.2006.05.005

Anglin, J. M. (1986). Semantic and Conceptual Knowledge Underlying the Child's Words. *Springer, New York, NY.*. https://doi.org/10.1007/978-1-4612-4844-6_4

Aristotle (1996). *Physics* (R. Waterfield & D. Bostock, Trans.). Oxford University Press. (Original work published ca. 350 B.C.).

Atran, S. (1994). *Core domains versus scientific theories: Evidence from systematics and Itza-Maya folkbiology* (pp. 316–340). Mapping the Mind: Domain Specificity in Cognition and Culture.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629
Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition, 194*, Article 104057. https://doi.org/10.1016/j.cognition.2019.104057
Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition, 167*, 25–37. https://doi.org/10.1016/j.cognition.2016.10.024
Bjorklund, D. F., & Thompson, B. E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology, 35*(2), 329–344. https://doi.org/10.1016/0022-0965(83)90086-3
Black, M. (1964). The gap between "Is" and "Ought". *The Philosophical Review, 73*(2), 165–181.
Blanchar, J. C., & Eidelman, S. (2021). Implications of Longevity Bias for Explaining, Evaluating, and Responding to Social Inequality. *Social Justice Research, 34*(1), 1–17. https://doi.org/10.1007/s11211-021-00364-1
Blancke, S., Schellens, T., Soetaert, R., Van Keer, H., & Braeckman, J. (2014). From ends to causes (and back again) by metaphor: The paradox of natural selection. *Science & Education, 23*(4), 793–808. https://doi.org/10.1007/s11191-013-9648-8
Borkenau, P. (1990). Traits as ideal-based and goal-derived social categories. *Journal of Personality and Social Psychology, 58*(3), 381–396. https://doi.org/10.1037/0022-3514.58.3.381
Brescoll, V., & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychological Science, 15*(8), 515–520. https://doi.org/10.1111/j.0956-7976.2004.00712.x
Burnett, R. C., Medin, D. L., Ross, N. O., & Blok, S. V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 59*(1), 3–10. https://doi.org/10.1037/h0087453
Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development, 1*(1), 37–41.
Carter, L. (2019). *Zak the Zigzagged Zebra.* https://www.assemblies.org.uk/pri/3329/stripes-and-zigzags.
Christensen, R. H. B. (2019). "ordinal—Regression Models for Ordinal Data". *R package version 2019.12-10*. https://CRAN.R-project.org/package=ordinal, 2019.
Coley, J. D., & Tanner, K. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE Life Sciences Education, 14*(1), ar8. https://doi.org/10.1187/cbe.14-06-0094
Darwall, S. (2013). Morality and principle. In D. Bakhurst, M. O. Little, & B. Hooker (Eds.), *Thinking about reasons: themes from the philosophy of Jonathan Dancy* (pp. 168–191). Oxford University Press.
Dar-Nimrod, I., Heine, S. J., Cheung, B. Y., & Schaller, M. (2011). Do scientific theories affect men's evaluations of sex crimes? *Aggressive Behavior, 37*(5), 440–449. https://doi.org/10.1002/ab.20401
Daston, L. (2014). The naturalistic fallacy is modern. *Isis, 105*(3), 579–587.
Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21*(2), 234–242. https://doi.org/10.1177/0956797609357712
Del Pinal, G., & Reuter, K. (2017). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science, 41*, 477–501. https://doi.org/10.1111/cogs.12456
Eidelman, S., Pattershall, J., & Crandall, C. S. (2010). Longer is better. *Journal of Experimental Social Psychology, 46*(6), 993–998. https://doi.org/10.1016/j.jesp.2010.07.008
Eidelman, S., & Crandall, C. S. (2012). Bias in favor of the status quo. *Social and Personality Psychology Compass, 6*(3), 270–281. https://doi.org/10.1111/j.1751-9004.2012.00427.x
Eidelman, S., Crandall, C. S., & Pattershall, J. (2009). The existence bias. *Journal of Personality and Social Psychology, 97*(5), 765. https://doi.org/10.1037/a0017058
Foster-Hanson, E., & Rhodes, M. (2019a). Is the most representative skunk the average or the stinkiest? Developmental changes in representations of biological categories. *Cognitive Psychology, 110*, 1–15. https://doi.org/10.1016/j.cogpsych.2018.12.004
Foster-Hanson, E., & Rhodes, M. (2019b). Normative Social Role Concepts in Early Childhood. *Cognitive Science, 43*(8), 1–18. https://doi.org/10.1111/cogs.12782
Foster-Hanson, E., & Rhodes, M. (2022, April 26). *Stereotypes as prototypes in children's gender concepts*. https://doi.org/10.31234/osf.io/ncxds
Foster-Hanson, E.*, Moty, K.*, Cardarelli, A., Ocampo, J. D., & Rhodes, M. (2020). Developmental changes in strategies for gathering evidence about biological kinds. *Cognitive Science.* https://doi.org/ 10.1111/cogs.12837.
Foster-Hanson, E., Roberts, S. O., Gelman, S. A., & Rhodes, M. (2021). Categories convey prescriptive information across domains and development. *Journal of Experimental Child Psychology, 212*, Article 105231. https://doi.org/10.1016/j.jecp.2021.105231
Friedrich, J., Kierniesky, N., & Cardon, L. (1989). Drawing moral inferences from descriptive science: The impact of attitudes on naturalistic fallacy errors. *Personality and Social Psychology Bulletin, 15*(3), 414–425. https://doi.org/10.1177/0146167289153011
Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual review of anthropology, 40*, 379–398. https://doi.org/10.1146/annurev-anthro-081309-145822
Gelman, S. A., & Waxman, S. R. (2009). Response to Sloutsky: Taking development seriously: Theories cannot emerge from associations alone. *Trends in cognitive sciences, 13*(8), 332. https://doi.org/10.1016/j.tics.2009.05.004
Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31*(2), 169–180. https://doi.org/10.3758/BF03194377
Goodwin, G. P., & Benforado, A. (2015). Judging the goring ox: Retribution directed toward animals. *Cognitive Science, 39*(3), 619–646. https://doi.org/10.1111/cogs.12175
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Mit Press.
Graves, J. L., Jr (2021). Human biological variation and the "normal". *American Journal of Human Biology, 33*(5), Article e23658. https://doi.org/10.1002/ajhb.23658
Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience, 59*(9), 792–799. https://doi.org/10.1525/bio.2009.59.9.10
Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9
Haward, P., Carey, S., & Prasada, S. (2021). The Formal Structure of Kind Representations. *Cognitive Science, 45*(10), Article e13040. https://doi.org/10.1111/cogs.13040
Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition, 176*, 255–268. https://doi.org/10.1016/j.cognition.2018.02.001
Hudson, W. D. (1969). *Is-Ought Question*. Springer.
Hume, D. (2003). *A treatise of human nature*. North Chelmsford, MA: Courier Corporation. Original work published 1739.
Ismail, I., Martens, A., Landau, M. J., Greenberg, J., & Weise, D. R. (2012). Exploring the effects of the naturalistic fallacy: Evidence that genetic explanations increase the acceptability of killing and male promiscuity. *Journal of Applied Social Psychology, 42*(3), 735–750. https://doi.org/10.1111/j.1559-1816.2011.00815.x
Järnefelt, E., Canfield, C. F., & Kelemen, D. (2015). The divided mind of a disbeliever: Intuitive beliefs about nature as purposefully created among different groups of non-religious adults. *Cognition, 140*, 72–88. https://doi.org/10.1016/j.cognition.2015.02.005
Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x
Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political psychology, 25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x
Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

Kay, A. C., Gaucher, D., Peach, J. M., Laurin, K., Friesen, J., Zanna, M. P., et al. (2009). Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be. *Journal of Personality and Social Psychology, 97*(3), 421–434. https://doi.org/10.1037/a0015997
Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In J. Tooby, A. M. Leslie, A. Caramazza, A. E. Hillis, E. C. Leek, M. Miozzo, … L. Cosmides (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). Cambridge University Press.
Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology, 35*(6), 1440–1452. https://doi.org/10.1037/0012-1649.35.6.1440
Kelemen, D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological Science, 15*(5), 295–301. https://doi.org/10.1111/j.0956-7976.2004.00672.x
Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution,* (pp. 66-92). Oxford: Oxford University Press. https://doi. org/10.1093/acprof:oso/9780199730421.003.0004.
Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*(1), 138–143. https://doi.org/10.1016/j.cognition.2009.01.001
Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General, 142*(4), 1074–1083. https://doi.org/10.1037/a0030399
Kim, S., & Murphy, G. L. (2011). Ideals and category typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1092–1112. https://doi.org/10.1037/a0023916
Kierniesky, N. C., & Sobus, M. (1989). The naturalistic fallacy: Moral inferences drawn from research with children versus adults. *Psychological Reports, 65*(2), 475–479. https://doi.org/10.2466/pr0.1989.65.2.475
Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*(4), 315–329. https://doi.org/10.1017/S0140525X10000907
Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition, 127*(2), 242–257. https://doi.org/10.1016/j.cognition.2013.01.005
Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition, 127*(2), 242–257. https://doi.org/10.1016/j.cognition.2013.01.005
Korman, J., & Khemlani, S. (2020). Teleological generics. *Cognition, 200*, Article 104157. https://doi.org/10.1016/j.cognition.2019.104157
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
Lagnado, D. A., Gerstenberg, T., & Zultan, R. I. (2013). Causal responsibility and counterfactuals. *Cognitive Science, 37*(6), 1036–1073. https://doi.org/10.1111/cogs.12054
Lane, M. (2020, September 9). Normative theory and the perils of normalization. [Paper for panel on "Things as They Should Be?" An Interdisciplinary Conversation in the Humanities]. 2020 Humanities Colloquium, Princeton, NJ, United States.
Lei, Ryan, et al. "How Race and Gender Shape the Development of Social Prototypes in the United States." PsyArXiv, 2 Feb. 2021. Web. *https://doi.org/10.31234/osf.io/yn8xs.*
Lemish, D., & Johnson, C. R. (2019). *The landscape of children's television in the US & Canada*. New York, NY: The Center for Scholars and Storytellers.
Lewry, C., Kelemen, D., & Lombrozo, T. (2021). Why belief in species purpose prompts moral condemnation of individuals who fail to fulfill that purpose. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43).
Liquin, E. G., & Lombrozo, T. (2018). Structure-function fit underlies the evaluation of teleological explanations. *Cognitive Psychology, 107*, 22–43. https://doi.org/10.1016/j.cogpsych.2018.09.001
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*(2), 167–204. https://doi.org/10.1016/j.cognition.2004.12.009
Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive Psychology, 65*(4), 457–485. https://doi.org/10.1016/j.cogpsych.2012.06.002
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*(11), 999–1006. https://doi.org/10.1111/j.1467-9280.2007.02015.x
Lombrozo, T., Shtulman, A., & Weisberg, M. (2006). The Intelligent Design Controversy: Lesson From Psychology and Education. *Trends in Cognitive Sciences, 10*(2), 56. https://doi.org/10.1016/j.tics.2005.12.001
Lombrozo, T. & Wilkenfeld, D. (2019). Mechanistic versus functional understanding. In Stephen Grimm (Ed.0, *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology* (pp. 209-229). Oxford University Press.
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*(1), 41–50. https://doi.org/10.3758/BF03211575
Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.
McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 123*(1/2), 125–148. http://www.jstor.org/stable/4321576.
Medin, D., & Atran, S. (1999). *Introduction. Folkbiology*. Cambridge, MA: MIT Press.
Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*(2), 242–279. https://doi.org/10.1016/0010-0285(87)90012-0
Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology, 12*(4), 496–522. https://doi.org/10.1016/0010-0285(80)90018-3
Mohr, R. D. (1977). Family resemblance, Platonism, universals. *Canadian Journal of Philosophy, 7*(3), 593–600. https://doi.org/10.2307/40230707
Moore, G. E. (2004). *Principia ethica*. Mineola, NY: Dover Publications. Original work published 1903.
Moore, R., Mitchell, G., Bally, R., Inglis, M., Day, J., & Jacobs, D. (2002). Undergraduates' understanding of evolution: Ascriptions of agency as a problem for student learning. *Journal of Biological Education, 36*(2), 65–71. https://doi.org/10.1080/00219266.2002.9655803
Murphy, G. L. (2002). *The big book of concepts* (p. 555). MIT Press.
Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 904. https://doi.org/10.1037/0278-7393.20.4.904
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289.
Nagel, T. (1979). Moral Luck. In *Mortal questions* (pp. 24–38). New York: Cambridge University Press.
Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 54. https://doi.org/10.1037/0278-7393.14.1.54
O'Connor, C. (2017). 'Appeals to nature' in marriage equality debates: A content analysis of newspaper and social media discourse. *British Journal of Social Psychology, 56*(3), 493–514. https://doi.org/10.1111/bjso.12191
Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185
Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences, 23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007
Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences of the United States of America, 114*(18), 4649–4654. https://doi.org/10.1073/pnas.1619717114
Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition, 145*, 30–42. https://doi.org/10.1016/j.cognition.2015.08.001
Plato, & Grube, G. M. A. (1974). *The Republic*. Indianapolis, IN: Hackett Publishing.

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*(1), 73–112. https://doi.org/10.1016/J.COGNITION.2005.01.003
Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33*(3), 401–448. https://doi.org/10.1111/j.1551-6709.2009.01018.x
Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology, 58*(6), 1048–1061. https://doi.org/10.1037/0022-3514.58.6.1048
Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1171. https://doi.org/10.1037/a0019765
Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory & Cognition, 38*(3), 377–388. https://doi.org/10.3758/MC.38.3.377
Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 665–681. https://doi.org/10.1016/S0022-5371(75)80055-7
Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science, 6*(1), 77–97. https://doi.org/10.1177/1745691610393525
Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic distance. *Journal of Verbal Learning and Verbal Behavior, 12*, 1–20.
Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science, 41*, 576–600. https://doi.org/10.1111/cogs.12443
Roberts, A. J., Handley, S. J., & Polito, V. (2021). The design stance, intentional stance, and teleological beliefs about biological and nonbiological natural entities. *Journal of Personality and Social Psychology, 120*(6), 1720. https://doi.org/10.1037/pspp0000383
Roberts, S. O., & Rizzo, M. T. (2021). The psychology of American racism. *American Psychologist, 76*(3), 475. https://doi.org/10.1037/amp0000642
Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9
Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2*(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491
Roseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02
Rosner, S. R., & Hayes, D. S. (1977). A developmental study of category item production. *Child Development, 48*(3), 1062. https://doi.org/10.2307/1128361
Rozin, P. (2005). The Meaning of "Natural": Process more important than content. *Psychological Science, 16*, 652–658. https://doi.org/10.1111/j.1467-9280.2005.01589.x
Rutjens, B. T., Sutton, R. M., & van der Lee, R. (2018). Not all skepticism is equal: Exploring the ideological antecedents of science acceptance and rejection. *Personality and Social Psychology Bulletin, 44*(3), 384–405. https://doi.org/10.1177/0146167217741314
Sánchez Tapia, I., Gelman, S. A., Hollander, M. A., Manczak, E. M., Mannheim, B., & Escalante, C. (2016). Development of teleological explanations in Peruvian Quechua-speaking and US English-speaking preschoolers and adults. *Child Development, 87*(3), 747–758. https://doi.org/10.1111/cdev.12497
Scott, S. E., & Rozin, P. (2020). Actually, natural is neutral. *Nature Human Behaviour, 4*(10), 989–990. https://doi.org/10.1038/s41562-020-0891-0
Shaw, E., & Blakey, R. (2020). Determinism, Moral Responsibility and Retribution. *Neuroethics, 13*, 99–113. https://doi.org/10.1007/s12152-019-09403-w
Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*(2), 170–194. https://doi.org/10.1016/j.cogpsych.2005.10.001
Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong.* New York: Basic Books.
Sloman, A. (1970). Ought' and 'better. *Mind, 79*(315), 385–394.
Strawson, P. F. (1962). *In Freedom and resentment and other essays (2008).* Routledge.
Tabb, K., Lebowitz, M. S., & Appelbaum, P. S. (2019). Behavioral genetics and attributions of moral responsibility. *Behavior Genetics, 49*(2), 128–135. https://doi.org/10.1007/s10519-018-9916-0
Tomasello, M. (2019). The moral psychology of obligation. *Behavioral and Brain Sciences., 1–33.* https://doi.org/10.1017/S0140525X19001742
Tomasello, M. (2020). The moral psychology of obligation. *Behavioral and Brain Sciences, 43*, 1–33.
Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science, 27*(8), 1109–1122. https://doi.org/10.1177/0956797616650875
Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology, 54*(9), 1735. https://doi.org/10.1037/dev0000555
Vasilyeva, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition, 204*, Article 104383. https://doi.org/10.1016/j.cognition.2020.104383
Vasilyeva, N., Wilkenfeld, D., & Lombrozo, T. (2017). Contextual utility affects the perceived quality of explanations. *Psychonomic Bulletin & Review, 24*(5), 1436–1450.
Von Fintel, K., & Iatridou, S. (2008). In *How to say ought in foreign: The composition of weak necessity modals* (pp. 115–141). Dordrecht: Springer.
Ware, E. A., & Gelman, S. A. (2014). You get what you need: An examination of purpose-based inheritance reasoning in undergraduates, preschoolers, and biological experts. *Cognitive Science, 38*(2), 197–243. https://doi.org/10.1111/cogs.12097
Wattenmaker, W. D. (1999). The influence of prior knowledge in intentional versus incidental concept learning. *Memory & cognition, 27*(4), 685–698.
Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13*(6), 258–263.
Weisberg, D. S., Landrum, A. R., Hamilton, J., & Weisberg, M. (2021). Knowledge about the nature of science increases public acceptance of science regardless of identity factors. *Public Understanding of Science, 30*(2), 120–138. https://doi.org/10.1177/0963662520977700
Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Ed.), *Handbook of child psychology: Cognition, perception, and language* (pp. 523–573). John Wiley & Sons Inc.
Wright, L. (1976). *Teleological explanations: An etiological analysis of goals and functions.* Berkeley, CA: University of California Press.
Wunderlich, C. M. (2020). *On the Permissibity of Abortion.* Wake Forest University.
Wysocki, T. (2020). Normality: A two-faced concept. *Review of Philosophy and Psychology, 11*(4), 689–716. https://doi.org/10.1007/s13164-020-00463-z
Ziff, P. (1972). *Understanding understanding.* Ithaca: Cornell University Press.