Explanation by essence

David Rose[1] and Shaun Nichols[2]
[1]Department of Psychology, Stanford University
[2]Department of Philosophy, Cornell University

Abstract

Lockean essentialism holds that people explain the characteristic features of natural kinds by appeal to hidden common causes—like DNA—that generate those features. Teleological essentialism, by contrast, proposes that people treat an entity's purpose as part of its essence and explain its features in terms of what it is for. Across three preregistered experiments, we tested these competing accounts by examining how adults explain why members of different categories have their characteristic features. In Experiment 1, participants explained features of artifacts and biological kinds; in both cases, they overwhelmingly appealed to purposes rather than underlying causes. Experiment 2 showed that people can and do appeal to causal mechanisms such as DNA when explaining within-species differences, yet they revert to purpose-based explanations when explaining between-species differences, which prompts species-level categorization. Experiment 3, using randomly selected biological and non-living natural kinds, found that participants again favored purposes for biological kinds but invoked causes for certain non-living kinds (e.g., lead, oxygen, platinum). Together, these findings suggest that for biological kinds, which are central to theories of psychological essentialism, the explanatory connection between essence and features is primarily understood in teleological, not causal, terms.

Keywords: categorization; teleology; essentialism; Lockean essentialism; explanation.

Corresponding author: David Rose (davdrose@stanford.edu). All the data, study materials, preregistrations, and analysis code are available here: `https://github.com/davdrose/explanation_by_essence`

Introduction

According to a prominent view of natural kinds, science sometimes uncovers the properties of things that make them what they are (Kripke, 1980; Putnam, 1975). The discovery of DNA is perhaps one of the most profound scientific discoveries in this regard, as it revealed a fundamental building block, a causal mechanism, that generates a wide array of biological kinds. The appeal to essences isn't relegated to science. Such essences are also thought to characterize people's beliefs about some categories (Gelman, 2003; Keil, 1989). For instance, adults think that a raccoon made to look like a skunk is still a raccoon, presumably because it retains the essence of raccoon over the transformation. That is, even though it superficially looks like a skunk, adults recognize that it still has the essence of raccoons. On the standard view in psychology—Lockean essentialism—adults associate that essence with an internal structure, e.g., raccoon DNA, that generates the observable features (e.g., Gelman, 2003; Keil, 1989; Medin & Ortony, 1989; Neufeld, 2022).

Lockean essentialism is an incredibly influential view about the psychology of categorization. It seems to characterize people's categorization judgments for a diverse range of kinds that include biological and social kinds (Atran, 1998; Gelman, 2003, 2013; Keil, 1989, 1994; Rothbart & Taylor, 1992). According to Lockean essentialism, while people may not always know what the underlying causal essence is, they can nonetheless represent essences as placeholders (Gelman, 2003). These placeholders might eventually be elaborated with the acquisition of scientific knowledge, like learning about DNA. Even without such elaboration, people nonetheless think that there is some underlying generative cause that will be discovered by science and that explains category-typical features of individuals.

A different proposal about the psychology of categorization is teleological essentialism. On this view, people treat what something is for—its purpose—as at least part of its essence. What makes something the kind of thing that it is is a function of its purpose. If a bee is made to look like a spider, so long as it preserves the purpose of bees, making honey and pollinating flowers, people think it is a bee (Rose & Nichols, 2019). But if the bee undergoes transformation and ends up with the purpose of spiders—spinning webs to catch and kill insects—people think it is no longer a bee.

Teleological essentialism contrasts with Lockean essentialism in a number of respects, including what people treat as the essence, which kinds are essentialized, and the relation between category features and essences. While research has focused on what people treat as the essence (e.g., Gelman, 2003; Joo & Yousif, 2022; Neufeld, 2021, 2022; Rose & Nichols, 2019; Toorman, 2023) and what kinds they treat as essentialized (e.g., Gelman, 2013; Rose & Nichols, 2020), here we focus on the relation between category features and essences.

Essences and features

On both Lockean and teleological essentialism, features are supposed to be explained by the category essence. However, these views characterize the explanatory relation differently. We begin with Lockean essentialism's characterization of the relation between features and essences before turning to teleological essentialism.

Lockean Essentialism

Observable features are represented as caused by the essence. This is a core part of Lockean essentialism. Indeed, this idea is suggested in Locke's (1690) own discussion of real essences. Real essences are, he writes, "the real internal, but generally (in substances) unknown constitution of things, *whereon their discoverable qualities depend*" (Book III, Chap III, section 15, emphasis added). In Locke's view, "all natural things...have a real, but unknown, constitution of their insensible parts; *from which flow those sensible qualities* which serve us to distinguish them one from another, according as we have occasion to rank them into sorts, under common denominations" (Book III, Chap III, section 17, emphasis added). He illustrates this with the example of gold, contrasting the "nominal essence" of gold with its "real essence":

> The nominal essence of gold is that complex idea the word gold stands for, let it be, for instance, a body yellow, of a certain weight, malleable, fusible, and fixed. But the real essence is the constitution of the insensible parts of that body, on which those qualities and all the other properties of gold depend. (Book III, c.6, section 2).

In contemporary cognitive science, the core idea is articulated in terms of generative causal models. For instance, Ahn (1999) writes, "when we think of a bird concept as a whole, *the most prominent structure is how the underlying essence would lead to surface features*, rather than how the surface features are related to each other into a single causal chain" (p. 1020, emphasis added). Leslie (2017) characterizes the central claim of psychological essentialism as follows: "We essentialize a kind if we form the (tacit) belief that there is some hidden, nonobvious, and persistent property or underlying nature shared by members of that kind *that causally grounds their common properties and dispositions*" (p. 405, emphasis added). Similarly, Neufeld (2022) writes:

> [P]sychological essentialism can be seen as a special case of the Causal Model Theory of concepts (Rehder, 2003a, 2003b, 2010, 2017), according to which concepts represent sets of features and causal relations between them. Psychological essentialism makes specific assumptions about some of the features and causal relations: essentialist concepts represent a common-cause structure, where the common cause is the category essence. (p. 2, fn. 2)

Neufeld uses a common cause graph to illustrate the general common cause model of the relation between category essences and features (Figure 1A). It's a generative causal model on which the characteristic features are explained by the common cause.

Likewise, Rehder (2007) notes that:

> [T]he mental act of categorization can be viewed as a case of causal reasoning in which properties like weight, body size, singing, and eating seeds provide inferential support for properties like bird DNA. This account provides an explanation for not only why people typically use observable features in classification, but also why they can override perceptual information in particular circumstances.

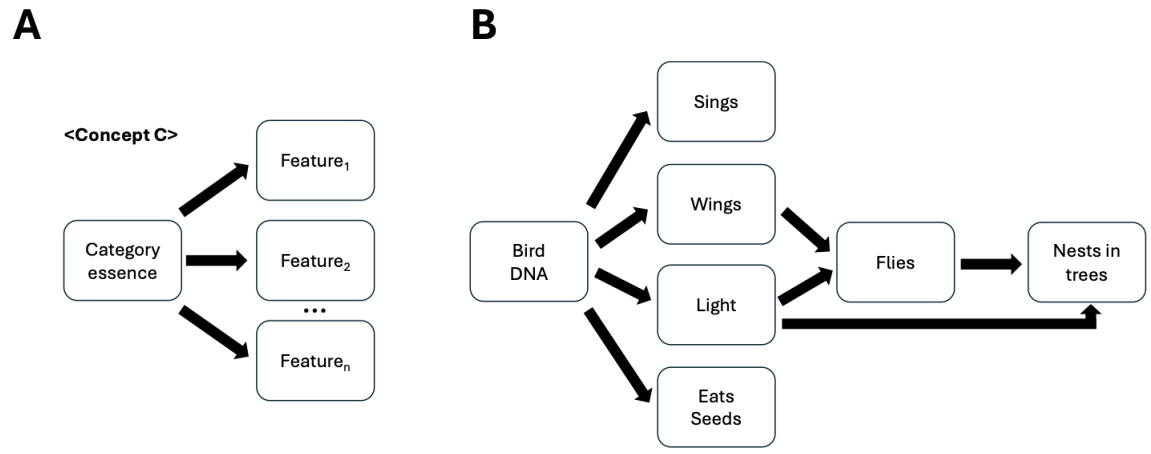

Figure 1

Common cause model*: (A) General common cause model of essence and features and (B) specific common cause model of essence and features for the bird category.*

> When one is told a story about how a particular bird's features are transformed through external intervention into those of an insect's, one recognizes that the underlying core properties remain unchanged and thus so does the animal's category membership (pp. 191–192).

Rehder's application of the common cause model to the particular case of the bird concept is reflected in Figure 1B.

In short, the Lockean essentialist envisions the relation between category essences and features in terms of a common cause model. Neufeld (2022) captures the crux of Lockean essentialism on this matter: "Concretely, the structure of essentialist concepts consists of the representation of an essence, a set of weighed typical features that are shared by category members, and *a representation of causal relations between the former and the latter*" (p. 2, emphasis added).

Teleological essentialism

According to teleological essentialism, the relation between a category essence and features is rather different from that proposed by Lockean essentialism. In the first place, the relation between a teleological essence and features cannot be captured by a common cause model. The telos doesn't directly generate the features. This looser relation between telos and features is familiar from everyday examples of the relation between goals and intentions. Intentions generate actions, and goals contribute to the formation of intentions. But the goal doesn't directly cause the intention, and this is partly because the goal often leaves latitude for different ways of satisfying the goal. I have the goal of getting a drink. This goal leads to my intention to go to the College bar, but the intention is influenced by lots of other things in addition to the goal—proximity, time of day, companions, and so forth. In like manner, if teleological essentialism is correct, the telos associated with a kind won't directly cause the features of the kind.

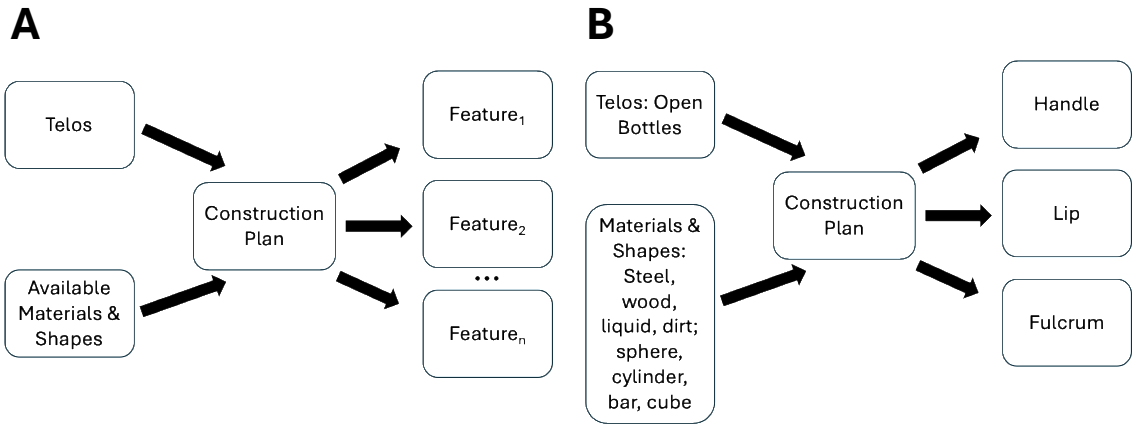

Figure 2

***Teleological model**: (A) Model of artifact generation and (B) specific model of artifact generation for a church key.*

In crucial respects, teleological essentialism treats species concepts as artifact concepts. This isn't to say that people must be thinking that there is an intentional designer for species. Instead, the crucial point is that the explanatory relation between features and essences is similar. We can thus understand that relation by focusing on how to model the generation of artifact features.

Artisans have particular tele in mind, and they know the available materials and options for manipulating the materials into various shapes. This feeds into a construction plan, which is effectively an intention for creating a particular object. The construction plan is to create a device that will serve the telos, given available materials (e.g., wood, steel, dirt), shapes (e.g., cube, cylinder, bar), and tools (e.g., forge, hammer, lathe). This requires assessing which materials and shapes would be good for the device. And there is typically no unique construction plan entailed by the telos and the available materials and shapes. We might depict this general structure graphically as in Figure 2A.

This model can be applied to the generation of a particular artifact, like a church key, as in Figure 2B. If one's goal is to devise something that can open bottles, many of the available materials won't work at all (e.g., liquids, dirt). Other materials can be preferentially ranked: steel is stronger and more rigid than wood, hence better as a lever. Similarly, some shapes (e.g., cube, sphere) won't work at all. This narrows the space, but one also needs to design a specific shape to enable the application of leverage to the bottle cap. The shape of a church key facilitates this.

As noted, there typically isn't a unique construction plan for creating an object to serve some telos. For the goal of opening bottles, church keys reflect one good plan, but wall-mounted bottle openers reflect another good plan. Similarly, the early artisans of forks and chopsticks had very similar goals in mind in crafting those artifacts, but the resulting objects have quite different features. Examples like this show that it's a mistake to think the telos directly causes the features of an artifact. The generation of an artifact depends on both the telos and a construction plan informed by available materials and shapes.

The importance of which materials and shapes are available also applies to evolu-

tionary adaptations. Evolution works with what it has, and this means that the adaptive solutions that evolution generates aren't uniquely specified by the adaptive problems abstractly characterized. The fangs of a snake and the spur of a platypus are different traits, but they have the same function—to deliver venom. Evolutionary adaptations also depend on what the organism already has in place. Although wheels on axles are great solutions for the problem of locomotion, no animals had structures in place that made this a possible adaptation.

The models of artifact generation (Figure 2 A & B) can be understood as real models of how artifacts are generated. But we think they can also be used to capture how ordinary people think about artifacts—in terms of goals and construction plans. More substantially, teleological essentialism holds that this model also captures how ordinary people think about essentialized categories like species.

Although tele are not the direct cause of the observable features, tele can partially explain why an object has the characteristic features it does. Again think of artifact generation (Figure 1 A & B). The construction plan is an intelligent process, and that is why the features of the output (i.e., the artifact) should make sense given the telos and the available materials and shapes. Thus, we can expect the features of an artifact to be conducive for achieving the function. Why do forks have tines? Because that makes forks good at conveying pieces of food to the mouth. In cognitive science, David Marr (1980) makes a similar point about cognitive adaptations:

> [B]ecause vision is used by different animals for such a wide variety of purposes, it is inconceivable that all seeing animals use the same representations; each can confidently be expected to use one or more representations that are nicely tailored to the owner's purposes (p. 32).

Something like this principle is operative in evolutionary thinking generally, already with Darwin's finches. The explanation for the large beak of *geospiza magnirostris* is that having a large beak is an adaptation that enables the bird to eat the large seeds that were in its environment. The size of the beak makes sense given the environment and the apparent purpose of the beak. We can explain features of an object by reflecting on the function of the object.

Teleological essentialism thus proposes that there is an informative link between essence and features. We can understand why members of a kind have characteristic features by appealing to the telos of that kind. As with Lockean essentialism, according to teleological essentialism, essences can be represented with placeholders, but they might often be elaborated with presumed tele. For instance, many people think that the function of bees is pollination (Rose & Nichols, 2019), and according to teleological essentialism, this will be part of the way the essence of bees is represented.

Experiment Overview

The models of category representation offered by Lockean and teleological essentialism both propose an explanatory link between essence and features. Moreover, while Figures 1 and 2 vastly oversimplify any real categories, both capture something important about the real world. The features of birds really are generated (partly) by DNA. And the

features of bottle openers really are made sense of in terms of construction plans informed by goals and effective materials. Our question is which of these models provides more insight into everyday categorization of essentialized kinds.

We designed experiments to investigate how people tend to explain characteristic features of a normal member of a kind. We can briefly state the theoretical commitments of each account on this issue:

> According to Lockean essentialism, people represent the essence as a common cause of the characteristic features of species, and thus can be expected to appeal to that underlying cause in explaining the features of species.
>
> According to teleological essentialism, people believe that species have tele, represent the telos of a species as part of the essence of the species, and thus can be expected to appeal to the telos when explaining the features of species.

If we assume that part of the way essences are understood by laypeople is in terms of the explanatory relation between the essence and characteristic features, then one way to examine these two accounts is by asking people to explain why members of the category have the characteristic features that they do, based on the essence that they have. That's what we did in Experiment 1, which examines how people think about the explanatory link between features and the purported species essence. In Experiment 2, we ask how people explain within- and between-species differences. Lastly, in Experiment 3, we examine a more representative set of items for both biological kinds and non-living natural kinds.

For all results reported in this paper, we analyzed the data using Bayesian logistic mixed effects models, using the default priors as specified by the `brms` package in R (Bürkner, 2017). More specifically, to test our hypotheses in each experiment, we examined the relative proportion of purpose to cause responses, fitting a Bayesian logistic regression model with condition as a predictor of responses and random intercepts for participants and items. We will refer to a statistical result of interest as "credible" when the 95% credible interval excludes 0 (or 1 when reporting odds ratios).

All experiments were pre-registered. The materials, data, pre-registrations, and analyses are available at `https://github.com/davdrose/explanation_by_essence`.

Experiment 1: Essence as Cause or Purpose for Artifacts and Biological Kinds

According to Lockean essentialism, people treat the essence of a kind—the true nature of a category that makes it what it is—as the underlying common cause that generates characteristic features of category members. Teleological essentialism is the view that people treat essences in terms of what they believe members of a category are for. According to teleological essentialism, an important part of the essence is the purpose that makes sense of the characteristic features.

On both views, part of the way essences are understood by laypeople is in terms of the explanatory relation between essence and characteristic features. To distinguish Lockean and teleological essentialism, we ask people to explain why members of the category have the characteristic features that they do, based on their essence (what makes them what they are).

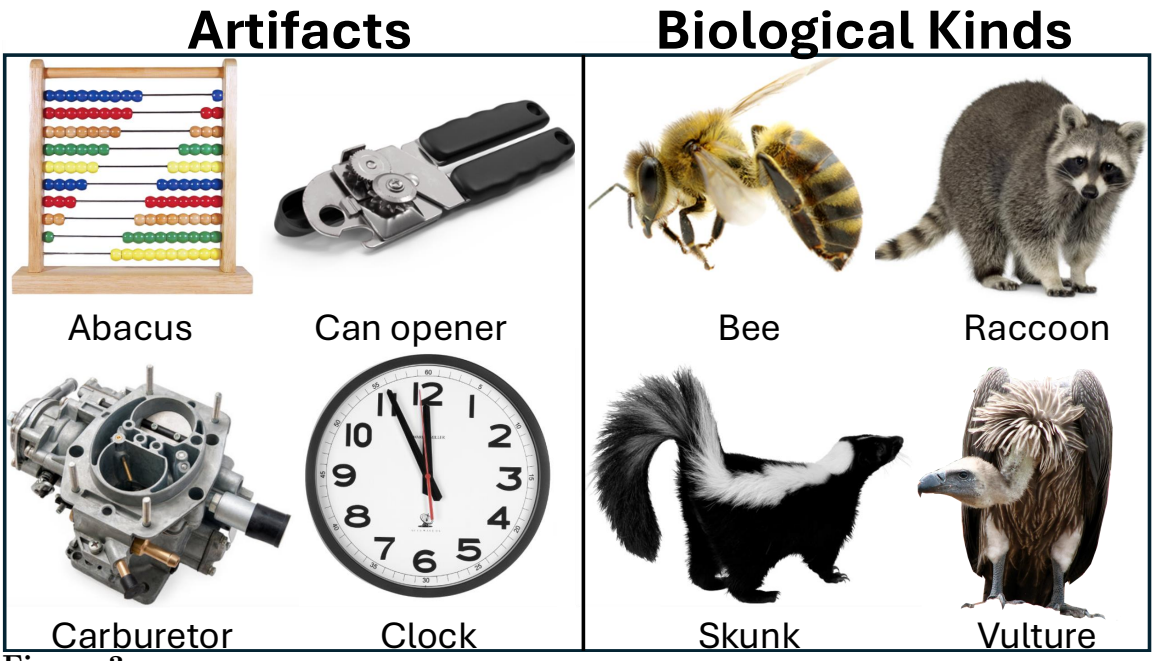

Figure 3

Experiment 1 materials*: In the artifact condition, people were given four artifacts and in the biological kind condition, they were given four biological kinds. Participants were first asked for each item what its most important features are that make it a member of its category. They were then asked when they think about what makes it a member of the category, what explains why it has those features.*

Methods

Participants

We recruited 150 participants (Age: $M = 42.69$, $SD = 12.93$; sex: 74 female, 74 male, 2 other/prefer not to answer; race: 1 American Indian/Alaskan Native, 9 Asian/Asian American, 11 Black/African American, 5 Latino/Hispanic, 122 White/European American, 2 other/prefer not to answer) through Prolific. Participants received compensation at a rate of $12 per hour.

Materials

The materials included biological kinds and artifacts. There were four artifacts and four biological kinds (Figure 3).

Procedure

Participants saw an image of a category member. Then they were asked two questions,"What do you think the most important features are that make a [category member] a [category member]?" and "When you think about what makes a [category member] a

Table 1

Sample participant responses with coding.

Response Coding	Participant Response
Purpose only	"To open a can for us." (*can opener, artifact condition*)
Cause only	"That's how the genetics of a typical raccoon are made up." (*raccoon, biological kind condition*)
Both purpose and cause	No examples (*in either the artifact or biological kind condition*)
Neither purpose nor cause	"The pronounced yellow and black coloring, often striped, is well known as what bees typically look like." (*bee, biological kind condition*)

[category member], what explains why it has those features?". Responses were made in text boxes.

Design

Participants were randomly assigned to either the artifact or biological kind condition. Within each condition, the order of the items was randomized. The questions were presented in a fixed order.

Results

We pre-registered coding responses as appealing to either purposes or causes (i.e., an underlying generative common cause). A response was coded as appealing to a purpose if it highlighted that a feature served a goal, such as survival, utility, or an ecological role. A response was coded as appealing to a cause if it highlighted a common cause that led to the feature. Responses that clearly referred to a purpose were coded as 1 for "purpose"; responses that clearly referred to a cause were coded as 1 for "cause". Responses that referred to neither a purpose or cause or were otherwise indeterminate were coded as 0 for "purpose" and 0 for "cause". Two researchers coded responses and agreed on 87.67% of responses. Disagreements were resolved through discussion. Table 1 shows examples of coded responses.

We had two pre-registered hypotheses. First, when asked to explain why artifacts have the features they do, people will be more inclined to appeal to purposes than to common causes. Second, when asked to explain why biological kinds have the features they do, people will be more inclined to appeal to purposes than to common causes (like DNA or genes).

The full set of response patterns is shown in Figure 4. In the artifact condition, 79% (*Confidence Interval*, CI [74%, 84%]) of responses appealed to a purpose only, while 0% (CI [0%, 0%]) of responses appealed to a cause only. As predicted, the estimated mean response in the artifact condition was credibly greater than 50% ($M = 100\%$, *Credible Interval*, CrI [99%, 100%]), indicating that people were more inclined to appeal only to a purpose relative to appealing only to a cause.

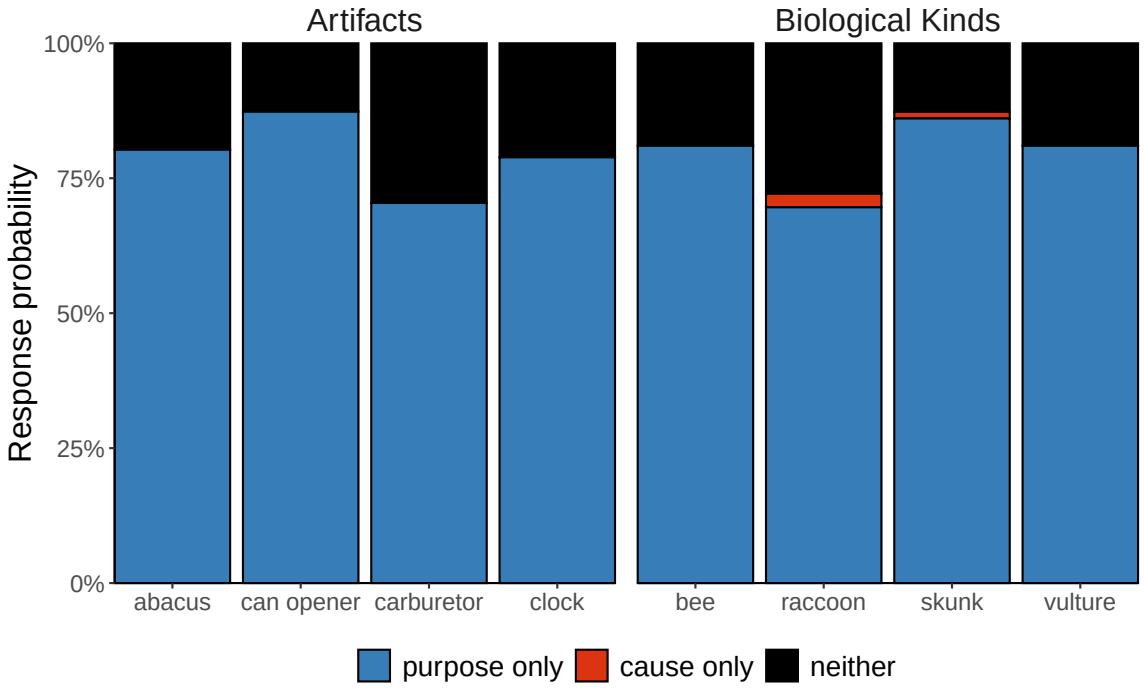

Figure 4

***Experiment 1 results**: Individual responses for each item in the artifact and biological kind conditions. "purpose only" means that only a purpose was appealed to in explaining the object's features; "cause only" means that only a common cause was appealed to in explaining the object's features; and "neither" indicates that neither a purpose nor a cause was appealed to. We also coded for whether responses appealed to "both" but there were none for either artifacts or biological kinds.*

Among the full response patterns in the biological kind condition, 79% (CI [75%, 84%]) of responses appealed to only a purpose, while 1% (CI [−0.13%, 2%]) of responses appealed to only a cause. Moreover, as predicted, the estimated mean response in the biological kind condition was credibly greater than 50% (M = 99%, CrI [99%, 100%]), indicating that people were more inclined to appeal only to a purpose relative to appealing only to a cause.

Though not pre-registered, we also found that there was no difference between artifacts and biological kinds in the extent to which people appealed to purposes. While the median odds ratio for the artifact/biological contrast was very large (OR = 6069), this point estimate is due to a ceiling effect, with probabilities for both conditions being near 1. The 95% credible interval was exceptionally wide and contained 1 (95% CrI [2.25×10^{-14}, 4.69×10^{20}]), indicating no credible evidence of a difference in odds between the artifact and biological kind conditions.

Discussion

We examined how people thought about the explanatory relation between the essence of a category and the characteristic features of members of that category. We attempted to elicit essentialist thinking by asking participants, "What do you think the most important features are that make a bee a bee?" and "When you think about what makes a bee a bee, what explains why it has those features?" As we predicted, participants tended to appeal to purposes in explaining why the category members had the features they did. They did this across a range of biological kinds and artifacts. This suggests that people do not represent the explanatory relation between a category's essence and its features as reflected in the common-cause model. Instead, their responses fit with the teleological model.

One response to our study is to maintain that when participants appeal to functions, they aren't thinking of the category's essence but rather something else. This response seems rather *ad hoc*, given that Lockean essentialism maintains that the essence is what explains the observable features. Moreover, the artifact results bolster the teleological essentialist interpretation. While Lockean essentialism maintains that people should appeal to underlying, generative common causes when explaining the features of biological kinds, they shouldn't do so when explaining the features of artifacts. Artifacts lack essences in the requisite sense, according to Lockean essentialism; there is no common cause that generates category-typical features to appeal to.

Teleological essentialism, by contrast, construes essences differently: people represent essences in terms of purposes. Thus, the relation between essences and features isn't causal but one of sense-making. And since artifacts are the central model of things that have purposes, people should appeal to purposes in explaining why objects have the features they do. Importantly, they should also do so for biological kinds. Our findings indicate that they do.

Still, it may be that people really represent the relation between essences and features as causal but don't appeal to DNA or genetics because they lack the requisite scientific knowledge that DNA in fact causes features. Lockean essentialists allow that people often represent essences with placeholders. If people are generally unfamiliar with DNA or genetics as causal mechanisms involved in the generation of features, then this would explain why they don't appeal to them when asked to explain the features of species. Our next study aims partly to determine whether people are actually unfamiliar with DNA or genetics as causal mechanisms involved in the generation of features.

Experiment 2: Within- and Between-Species Explanations

When people are asked to explain why members of a category have the features they do, we found that they tend to appeal to purposes, which is exactly what teleological essentialism predicts. Yet people may not generally appeal to Lockean essences like DNA or genetics because they are not familiar with these as causal mechanisms. However, we suspect that adults in our sample are familiar with DNA or genetics as causal mechanisms and that this will be reflected in explanations of certain within-species differences, such as differences between siblings. For example, when asked, "What explains why some chicks in a litter have larger beaks than others?", we anticipate that people will be inclined to

appeal to genetic differences. We use this *within-species* condition to confirm that people are at least familiar with DNA and genetics as causal mechanisms and can rely on them to explain some differences.

In another condition, we ask about differences *between-species*. We expect questions about differences between species to elicit categorical thinking about an essentialized category. By making salient the contrast between species, it is plausible that this would make the category essence salient. And given the essentialist models under consideration, this suggests that the category essence will be invoked in explaining the features. That is, if the category essence explains characteristic features of species members, then the link between essence and features is plausibly implicated when we invoke comparisons between species.

The critical question is whether a facility with genetic explanation will also be invoked in between-species cases where essentialist thinking is likely to occur. That is, if instead of asking, "What explains why some chicks in a litter have larger beaks than others?", we ask, "What explains why herons have larger beaks than robins?", will people still appeal to genetics? Insofar as the between-species question elicits essentialist thinking, Lockean essentialism predicts that people will explain differences in features between category members by appealing to causal mechanisms such as DNA or genes. Teleological essentialism, by contrast, predicts that people will explain differences between category members by appealing to a purpose.

Methods

Participants

We recruited 153 participants (Age: $M = 44.45$, $SD = 12.67$; sex: 74 female, 75 male, 4 other/prefer not to answer; race: 1 American Indian/Alaskan Native, 14 Asian/Asian American, 9 Black/African American, 5 Latino/Hispanic, 120 White/European American, 4 other/prefer not to answer) through Prolific. Participants received compensation at a rate of $12 per hour.

Materials

The materials involved biological kinds where siblings' features were described or a feature was compared between two animals. The full set of materials is shown in Table 2.

Procedure

Participants read a description of the animal and its feature and were then asked why they thought the animals had these different features (see Table 2). Responses were made in text boxes.

Design

Participants were randomly assigned to either the within- or between-biological kind condition. In each condition, the order of the items was randomized.

Table 2

Experiment 2 materials. There were two conditions: a within-species and a between-species condition. In each, participants were given four descriptions of differences and then asked to explain those differences. Responses were written in a text box.

Condition	Description	Question
Within	In a litter of herons, some chicks have larger beaks than others.	What do you think explains why some chicks in a litter have larger beaks than others?
	In a litter of polar bears, some cubs have whiter fur than others.	What do you think explains why some polar bears have whiter fur than others?
	In a litter of rabbits, some babies have hairier toes than others.	What do you think explains why some rabbits have hairier toes than others?
	In a litter of wolves, some pups have sharper teeth than others.	What do you think explains why some wolves have sharper teeth than others?
Between	Herons have larger beaks than robins.	What do you think explains why herons have larger beaks than robins?
	Polar bears have whiter fur than grizzly bears.	What do you think explains why polar bears have whiter fur than grizzly bears?
	Rabbits have hairier toes than raccoons.	What do you think explains why rabbits have hairier toes than raccoons?
	Wolves have sharper teeth than groundhogs.	What do you think explains why wolves have sharper teeth than groundhogs?

Results

We pre-registered coding responses as appealing to either purposes or causes (i.e., an underlying generative common cause), using the same scheme as in Experiment 1. Two researchers coded responses and agreed on 89.54% of responses. Disagreements were resolved through discussion. Table 3 shows examples of coded responses.

We had two pre-registered hypotheses. First, when asked to explain why different individuals within a biological kind—in particular, siblings—have different features, people will be more inclined to appeal to causal mechanisms (like DNA or genes) than purposes. Second, when asked to explain why members of different biological kinds have different features, people will be more inclined to appeal to purposes than causal mechanisms (like DNA or genes).

The full set of response patterns is shown in Figure 5. In the within-biological kind condition, 62% (CI [57%, 68%]) of responses appealed only to a cause, while 3% (CI [1%, 6%]) of responses appealed only to a purpose. As predicted, the estimated mean response in the within-biological kind condition was credibly lower than 50% ($M = 0\%$, CrI [0.001%, 0%]), indicating that people were more inclined to appeal only to a cause relative

Table 3

Sample participant responses with coding in the between-species condition.

Response Coding	Participant Response
Purpose only	"Herons need larger beaks for eating fish." (*heron, between-species*)
Cause only	"They genetically have thicker fur than raccoons." (*rabbits, between-species*)
Both purpose and cause	"Their fur is a genetic adaptation to their environment. Having white fur gives polar bears camouflage in snowy conditions, while darker fur helps grizzly bears blend in the woods." (*polar bears, between-species*)
Neither purpose nor cause	"Wolves eat and kill big animals." (*wolves, between-species*)

to appealing only to a purpose.

Of the full response patterns in the between-biological kind condition, 79% (CI [75%, 84%]) of responses appealed only to a purpose, while 1% (CI [0.03%, 2%]) of responses appealed only to a cause. Moreover, as predicted, the estimated mean response in the between-biological kind condition was credibly greater than 50% (M = 100%, CrI [99%, 100%]), indicating that people were more inclined to appeal only to a purpose relative to appealing only to a cause.

Discussion

We found that when asked about within-species differences, such as "What explains why some chicks in a litter have larger beaks than others?", people tended to appeal to DNA or genetics. This suggests that people are at least familiar with DNA and genetics as causal mechanisms involved in generating features, and that they can and do appeal to them in explaining some features. Importantly, when we instead ask about category differences, such as "What explains why herons have larger beaks than robins?", people no longer appeal to DNA or genetics.

Instead, they largely appeal to purposes. This indicates that people are indeed familiar with and capable of appealing to DNA or genetics in some explanations of features, but when it comes to explanations of differences between species—where essences are most likely to be prominent—people largely draw on purposes to explain the link between features and essences.

There are two important limitations to our experiments so far. First, the items were hand-picked. Second, we have only focused on biological kinds and artifacts. But non-living natural kinds—such as gold and water—are also thought to be essentialized (see Gelman, 2003; Keil, 1989; but see Malt, 1994). It may be that people appeal to Lockean essences—like atomic number 79 or H_2O—when explaining why some natural kinds have the features that they do. Our third study addresses both limitations.

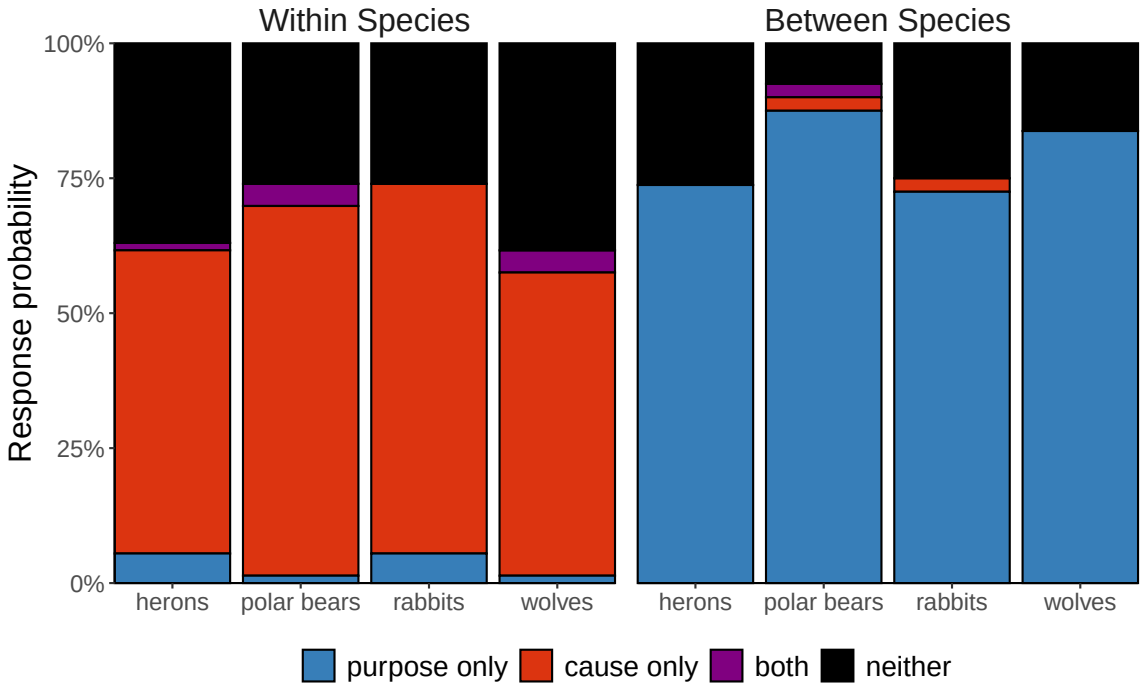

Figure 5

***Experiment 2 results**: Individual responses for each item in the within- and between-species conditions. "purpose only" means that only a purpose was appealed to in explaining the features; "cause only" means that only a common cause was appealed to in explaining the features; "both" means that both a purpose and cause were appealed to and "neither" indicates that neither a purpose nor a cause was appealed to.*

Experiment 3: Essence as Cause or Purpose for Biological Kinds and Non-Living Natural Kinds

Methods

Participants

We recruited 305 participants (Age: $M = 44.64$, $SD = 13.71$; sex: 148 female, 155 male, 2 other/prefer not to answer; race: 4 American Indian/Alaskan Native, 21 Asian/Asian American, 37 Black/African American, 18 Latino/Hispanic, 220 White/European American, 5 other/prefer not to answer) through Prolific. Participants received compensation at a rate of $12 per hour.

Materials

The materials involved biological kinds and non-living natural kinds. The items were generated by taking all nouns from `https://www.desiquintans.com/downloads/nounlist/nounlist.txt` and then having Claude tag all biological and non-living natural

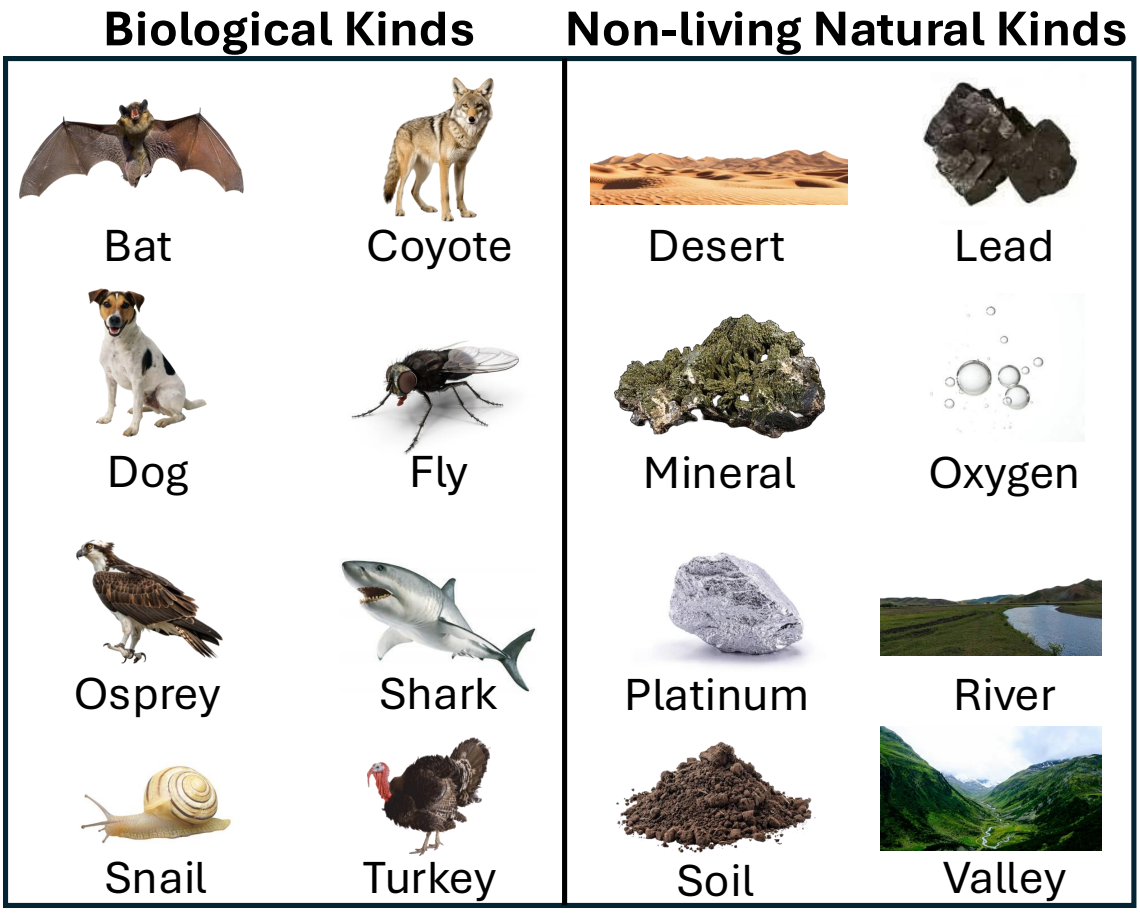

Figure 6

Experiment 3 materials*: In the biological kind condition, people were given eight biological kinds and in the non-living natural kind condition, they were given eight non-living natural kinds. Participants were first asked for each item what its most important features are that make it a member of its category. They were then asked when they think about what makes it a member of the category, what explains why it has those features.*

kinds. From that list, we used `R` to randomly select eight items for each kind. The materials are shown in Figure 6.

Procedure

The procedure was the same as in Experiment 1.

Design

The design was the same as in Experiment 1, except that the four items participants received were randomly selected from among the set of eight.

Table 4

Sample participant responses with coding.

Response Coding	Participant Response
Purpose only	"The wings help it to fly and its teeth help it to eat." (*bat, biological kind condition*)
Cause only	"Chemical makeup structure of the molecule." (*platinum, non-living natural kind condition*)
Both purpose and cause	"Its genetics are what makes it into a shark. It has those features to make it into a ferocious ocean hunter." (*shark, biological kind condition*)
Neither purpose nor cause	"It exists due to its location—a setting within a landscape." (*valley, non-living natural kind condition*)

Results

We pre-registered coding responses as appealing to either purposes or causes (i.e., an underlying generative common cause), using the same scheme as in Experiment 1. Two researchers coded responses and agreed on 84.84% of responses. Disagreements were resolved through discussion. Table 4 shows examples of coded responses.

We had one pre-registered hypothesis. When asked to explain why biological kinds have the features they do, we predicted that people would be more inclined to appeal to purposes than common causes (like DNA or genes). We did not have any predictions for the non-living natural kinds condition.

The full set of response patterns is shown in Figure 7. In the biological kind condition, 67% (CI [64%, 71%]) of responses appealed only to a purpose, while 2% (CI [1%, 3%]) of responses appealed only to a cause. As predicted, the estimated mean response in the biological kind condition was credibly greater than 50% ($M = 100\%$, CrI [99%, 100%]), indicating that people were more inclined to appeal only to purposes relative to appealing only to a cause.

Among the full response patterns in the non-living natural kind condition, 5% (CI [3%, 7%]) of responses appealed only to a purpose, while 25% (CI [21%, 28%]) of responses appealed only to a cause. Some of the items, such as lead, minerals, oxygen, and platinum, had higher rates of people appealing only to causes, with lead being the highest at 48%. The estimated mean response in the non-living natural kind condition was credibly lower than 50% ($M = 0\%$, CrI [0%, 1%]), indicating that people were more inclined to appeal only to causes relative to appealing only to a purpose.

Discussion

We examined whether people are more inclined to appeal to Lockean or teleological essences when explaining the features of non-living natural kinds and biological kinds, among a randomly selected sample of items from each category. Among a randomly selected set of biological kinds, people were more inclined to explain the features of category members by appealing to purposes. But for non-living natural kinds, people rarely appealed to purposes.

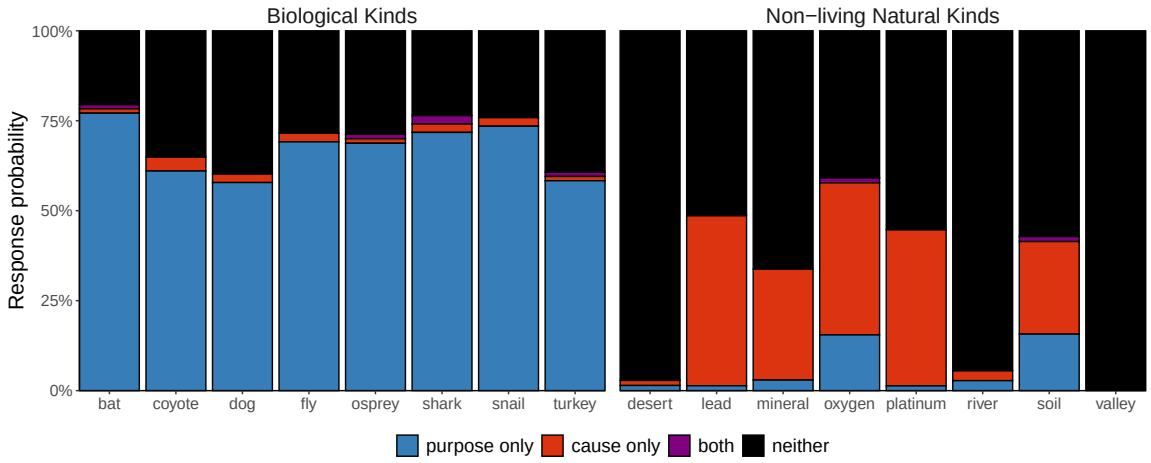

Figure 7

***Experiment 3 results**: Individual responses for each item in the biological kind and non-living natural kind conditions. "purpose only" means that only a purpose was appealed to in explaining the object's features; "cause only" means that only a common cause was appealed to in explaining the object's features; "both" indicates that they appealed to both a purpose and a cause; and "neither" indicates that neither a purpose nor a cause was appealed to.*

While for many non-living natural kinds people were largely inclined to appeal to neither a purpose nor a cause, they were nonetheless more inclined to appeal to causes than to purposes. And at least for some non-living natural kinds—such as lead, oxygen, and platinum—upward of 50% of participants were inclined to explain their features by appealing to causes.

General Discussion

What makes something the kind of thing that it is? Lockean essentialism maintains that, at least for natural kinds, such as biological and non-living natural kinds, people believe that what makes them what they are is an underlying common cause, and that these common causes might be revealed by science. In contrast, teleological essentialism maintains that for both natural kinds, like biological kinds, and artifacts, people believe that what makes them what they are is a function of what they are for—their purpose. Both Lockean and teleological essentialism differ in a number of important respects on the psychology of categorization, including what people treat as the essence, the kinds that people believe are essentialized, and the relation between category features and essences. Here we focused on how people represent the relation between category features and essences.

Lockean essentialism maintains that people represent the relation between category essences and features in terms of an underlying common cause. Teleological essentialism treats the relation between category essences and features differently. Purposes do not cause features. Instead, purposes make sense of the features that a kind tends to have. Lockean and teleological essentialism thus make different predictions about what people will appeal to when asked to explain why category members have the features they do.

Lockean essentialism predicts that people will appeal to underlying causes, like DNA or genetics; teleological essentialism predicts that people will appeal to purposes. We tested this in three experiments.

When asked, in Experiment 1, "What do you think the most important features are that make a bee a bee?" and "When you think about what makes a bee a bee, what explains why it has those features?", we found that people tended to appeal to purposes in explaining why the category members had the features they did. They did this for all of the biological items: bees, vultures, raccoons, and skunks. Importantly, participants made very similar kinds of purpose-based judgments when asked to explain why a range of artifacts (abacuses, carburetors, can openers, and clocks) have the features they do. This suggests a close parallel between how people represent biological and artifactual categories. They do not explain the features of biological or artifactual kinds by appealing to common causes. Instead, for both kinds of entities, they appeal to purposes.

Perhaps our participants do not appeal to DNA or genes to explain features because they are generally not aware of them as causal mechanisms. Lockean essentialism holds that people often represent essences as placeholders and rarely elaborate them with scientific knowledge. In Experiment 2, we explored whether participants can appeal to DNA or genes in explaining some features. We found that when asked about within-species differences, such as "What explains why some chicks in a litter have larger beaks than others?", people tended to appeal to DNA or genetics. This suggests that our participants are indeed familiar with DNA or genetics as causal mechanisms involved in generating features. Importantly, however, when we asked about category differences, such as "What explains why herons have larger beaks than robins?", people no longer appealed to DNA or genetics. Instead, they largely appealed to purposes.

Experiments 1 and 2 included items that we selected. In Experiment 3, we randomly generated items for the study. In addition to randomly generating biological kinds, we also randomly generated non-living natural kinds. In the literature on psychological essentialism, natural kinds in general are thought to be represented in terms of a common-causal essence, and some of the standard examples, like gold and water, are non-living natural kinds. We found that even among a randomly selected set of biological kinds, people were inclined to explain their features by appealing to purposes. This provides stronger evidence for the general claim that people think of species essences in terms of teleology. For non-living natural kinds, people tended not to appeal to either purposes or common causes. However, people were more likely to appeal to causes than to purposes. And for some non-living kinds—like lead, oxygen, and platinum—people were more likely to explain their features by appealing to common causes.

Together, our findings suggest that people do not generally represent the relation between essences and features as Lockean essentialism envisions. While it is standardly held that the relation between features and essences is a causal relation, where the essence is the common cause of category-typical features, people do not appeal to underlying common causes like DNA or genes when explaining why species like bees have the features they do. People appeal to their purposes. These are not represented as simple generative causes. Rather, they are treated as what partially explains and makes sense of the features that category members tend to have. This supports the idea that people treat essences in terms of purposes, just as teleological essentialism maintains.

At the same time, there are some kinds for which people do seem to appeal to Lockean essences: certain non-living natural kinds. Interestingly, unlike biological kinds, where people appeal to purposes across a wide range of kinds, among non-living natural kinds there is a much more limited subset of kinds that are explained in terms of Lockean essences. These include kinds like lead, oxygen, and platinum. So there may be some more circumscribed role for Lockean essentialism in people's representation of categories. At the same time, we treated appeals to chemical composition as appeals to underlying common causes because proponents of Lockean essentialism treat them as such (e.g., Gelman, 2003). For instance, much like DNA, having atomic number 79 or being composed of H_2O is treated as a common cause of category-typical features. But atomic number 79 and H_2O are not obviously common causes of features. Does H_2O cause a liquid to be clear, odorless, and tasteless? If Lockean essentialism has it that people represent essences as underlying common causes of observable features, then it is not entirely clear that it even characterizes non-living natural kinds like water and gold, or lead, oxygen, and platinum.

However people view natural kinds, it is clear that people can and do sometimes use DNA to categorize (e.g., Dar-Nimrod & Heine, 2011). Our results from Experiment 2, where we found that people explain within-species differences by appealing to DNA, show that people are familiar with DNA. But do people use DNA to categorize because they treat it as a common cause of category-typical features? It may be that when using DNA to inform some of their categorization decisions, people treat DNA like a fingerprint—a sort of unique identifier. Whatever the case may be, at least for some kinds that are of central interest in discussions of psychological essentialism—biological kinds—teleological essentialism better captures how people represent the explanatory structure of these kinds.

Prior work on teleology and categorization found that functional features are judged more diagnostic of biological kind membership because they are seen as stable across time and grounded in evolutionary history (Lombrozo & Rehder, 2012). These findings are of course consistent with ours, but our findings point to a deeper source of the privilege of purpose. Teleology plays a role in categorization because people represent functions as part of what something is—an essential rather than merely historical or statistical property. At the same time, we recognize that some maintain that essentialism itself mischaracterizes how people represent categories (e.g., Strevens, 2000) or that teleology should not be regarded as part of essence representation (e.g., Kelemen & Carey, 2007). Yet even if one rejects essentialism altogether, our results remain striking. They show that teleological thinking is deeply woven into categorization, shaping how people understand kind membership. This pattern is evident not only in the present studies but also in work on transformation and "switched-at-birth" tasks (Rose & Nichols, 2019). Whether or not one accepts the essentialist framework, appealing to teleology seems to provide a richer and more illuminating account of category representation than the common-cause models central to Lockean essentialism (e.g., Gelman, 2003; Keil, 1989; Neufeld, 2022). Still, the dominant view is that some form of essentialism correctly captures how people represent some categories (e.g., Ahn et al., 2001; Haslam, Rothschild, & Ernst, 2000; Rhodes & Gelman, 2009). We have followed that prevailing consensus. But we add that the kind of essentialism that best captures how people represent some categories, including species, is teleological essentialism.

One virtue of our studies is that they did not impose forced choices over a limited option set. Rather, our studies allowed for open responses. People spontaneously gener-

ated teleological explanations for traits. Among other things, this shows the fecundity of folk teleological reasoning. Indeed, our explanatory tasks bring out a limitation of Lockean essentialism that has rarely been registered. Although Lockean essentialism analyzes essentialized categories as generative models, the generative models are very weak. Lockean essentialism does not motivate any distinctive predictions from essences. This is obviously the case if one has a pure placeholder for the category—a placeholder essence can provide no distinctive predictions or explanations regarding the features that the essence generates. But even if one moves from a pure placeholder essence to the more scientifically elaborated view that the essence of birds is bird DNA, this still does not provide any new explanatory power unless one knows a great deal of molecular genetics. In particular, it does not help one make any distinctive predictions or explanations regarding the features of birds. All one knows is that the bird DNA causes the features that one already knows about. In this respect, if categories are represented as Lockean essences, the explanatory utility is rather thin. By contrast, insofar as one has even a glimmer of an idea about the purpose of an organism, this can be more generative of explanations. And this is exactly what we see in our studies. Thus, not only do our findings confirm the predictions of teleological essentialism, they also reveal why teleological reasoning carries greater explanatory potential.

Our findings also bear on theories of explanation more broadly. Contemporary accounts distinguish mechanistic explanations, which appeal to underlying causal organization and mechanisms (e.g., Lombrozo, 2012; Machamer, Darden, & Craver, 2000), from teleological explanations, which appeal to goals or functions (e.g., Kelemen, 1999; Lombrozo & Rehder, 2012). Prior work has largely treated these as alternative explanatory modes that people flexibly deploy to make sense of the same phenomena. Our results suggest something deeper: the explanatory mode people favor reflects how they represent what something is. For biological kinds, purposes are not merely convenient explanations for features—they are part of what people take the essence to be. Teleological and mechanistic explanation thus differ not only in form but in ontological depth: they track different ways of understanding the nature of things. This suggests that explanation and categorization are deeply intertwined—explanatory forms reflect the underlying conceptual structures that people use to understand kinds. Teleological explanation, in particular, appears to provide the intuitive framework through which people represent essences and make sense of characteristic features, whereas mechanistic explanation—though central to scientific reasoning—plays a more constrained role in everyday conceptual understanding.

There is, at the same time, great appeal to Lockean essentialism. It draws from the mechanistic framework that is endorsed by most contemporary people engaged in scientific work. Mechanistic thinking dominates how scientists think about the world (and we include ourselves here). But it is a bold hypothesis that this mechanistic framework captures the core of how ordinary people think about natural categories like species. Purposes are pervasive in ordinary thinking (e.g., Kelemen, 1999), including thinking about natural categories (e.g., Rose & Nichols, 2020). And purposes generate a very different explanatory profile than mechanisms. Treating science as a model for how people learn about and represent the world (e.g., Gopnik, 1996) has profoundly changed how we understand human cognition and development. It is easy to appreciate its allure when it comes to characterizing how people represent categories. But on this matter at least, the mechanistic worldview that came in the wake of the scientific revolution has yet to displace the Aristotelian, teleological

view that is built into the ordinary representation of categories.

Author contributions

Conceptualization: DR & SN; Methodology: DR & SN; Software: DR; Validation: DR; Formal Analysis: DR; Investigation: DR & SN; Data Curation: DR; Writing—Original Draft: DR; Writing—Review & Editing: DR & SN; Visualization: DR; Supervision: DR & SN; Project Administration: DR & SN; Funding Acquisition: SN

Acknowledgments

We thank Eleanor Neufeld, Lance Rips and Jeske Toorman for their feedback and discussion.

References

Ahn, W.-k. (1999). Effect of causal structure on category construction. *Memory & Cognition*, *27*(6), 1008–1023. doi: 10.3758/BF03201229

Ahn, W.-k., Kalish, C. W., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., ... Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition*, *82*(1), 59–69. doi: 10.1016/S0010-0277(01)00165-9

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, *21*(4), 547–569. doi: 10.1017/S0140525X98001277

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive persistence of genetic ideas. *Psychological Bulletin*, *137*(5), 800–818. doi: 10.1037/a0021860

Gelman, S. (2003). *The essential child*. Oxford University Press.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195154061.001.0001

Gelman, S. A. (2013). Artifacts and essentialism. *Review of Philosophy and Psychology*, *4*, 449–463. doi: 10.1007/s13164-013-0142-7

Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, *63*(4), 485–514. doi: 10.1086/289971

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, *39*(1), 113–127. doi: 10.1348/014466600164363

Joo, S.-Y., & Yousif, S. R. (2022). Are we teleologically essentialist? *Cognitive Science*, *46*(11), e13202. doi: 10.1111/cogs.13202

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

Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition*, *70*(3), 241–272. doi: 10.1016/S0010-0277(99)00010-4

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation* (pp. 212–230). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199280698.003.0011

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Leslie, S.-J. (2017). The original sin of cognition. *The Journal of Philosophy*, *114*(8), 395–421. doi: 10.5840/jphil2017114827

Locke, J. (1690). *An essay concerning human understanding*. London: Thomas Basset.

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 260–276). Oxford: Oxford University Press.

Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive*

Psychology, *65*(4), 457–485. doi: 10.1016/j.cogpsych.2012.06.002

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. doi: 10.1086/392759

Malt, B. C. (1994). Water is not h2o. *Cognitive Psychology*, *27*(1), 41–70. doi: 10.1006/cogp.1994.1012

Marr, D. (1980). Visual information processing: The structure and creation of visual representations. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 199–218. doi: 10.1098/rstb.1980.0092

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511529863.009

Neufeld, E. (2021). Against teleological essentialism. *Cognitive Science*, *45*(4), e12961. doi: 10.1111/cogs.12961

Neufeld, E. (2022). Psychological essentialism and the structure of concepts. *Philosophy Compass*, *17*(5), e12823. doi: 10.1111/phc3.12823

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science, vol. 7* (pp. 215–271). Minneapolis: University of Minnesota Press.

Rehder, B. (2003a). A causal-model theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 734–747. doi: 10.1037/0278-7393.29.4.734

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1141–1159. doi: 10.1037/0278-7393.29.6.1141

Rehder, B. (2007). Essentialism as a generative theory of classification. In W.-k. Ahn, R. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). New York: Oxford University Press.

Rehder, B. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1171–1206. doi: 10.1037/a0020132

Rehder, B. (2017). The psychology of causal-based categorization. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 347–370). Oxford: Oxford University Press.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274. doi: 10.1016/j.cogpsych.2009.05.001

Rose, D., & Nichols, S. (2019). Teleological essentialism. *Cognitive Science*, *43*(4), e12725. doi: 10.1111/cogs.12725

Rose, D., & Nichols, S. (2020). Teleological essentialism: Generalized. *Cognitive Science*, *44*(3), e12818. doi: 10.1111/cogs.12818

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). Newbury Park, CA: Sage Publications.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition*, *74*(2), 149–175.

doi: 10.1016/S0010-0277(99)00072-6

Toorman, J. (2023). Against arguments from diagnostic reasoning. *Cognitive Science*, *47*(11), e13376. doi: 10.1111/cogs.13376