Teleology Beyond Diagnosticity: Evidence for Teleological Essentialism

David Rose[1], Shaun Nichols[2], and Tobias Gerstenberg[1]
[1]Stanford University
[2]Cornell University

Abstract

Teleological essentialism proposes that people treat purposes as part of a kind's essence. This contrasts with standard psychological essentialism, where essences are hidden causes that are discovered through science (e.g., DNA). Recent critiques of teleological essentialism argue that manipulating a thing's purpose affects people's categorization judgments only indirectly via diagnostic inferences about the underlying hidden causes. Here, we demonstrate that the effect of teleology on categorization cannot be explained by such diagnostic reasoning alone. In Experiment 1, a bee transformed to look like a spider—retaining bee DNA but acquiring a spider's purpose—is more likely to be categorized as a spider. In Experiment 2, when DNA is explicitly framed as the common cause of both a superficial feature and a purposeful feature, altering the purpose affects categorization judgments more than altering a superficial feature, even when the DNA is held fixed. Together, our findings support teleological essentialism and show that purpose plays a special role in categorization, one that cannot be explained merely in terms of diagnostic inference.

Keywords: teleology; psychological essentialism; categorization; diagnostic reasoning; causal interventions

Introduction

How do we determine whether something is a bee? Here is one answer: if it looks like a bee, then it's a bee. But suppose that some scientists perform an operation on a bee. They remove its wings and antennae, lengthen its legs and add a new pair of legs. After the procedure, it looks like a spider. What is it after the procedure? Is it still a bee or is it now a spider? If what it looked like was all that mattered, it would now be a spider. Yet people judge that it's still a bee (Keil, 1989).

People sometimes categorize things by their essence (e.g., Medin & Ortony, 1989)—a thing's true nature which cannot be directly observed yet gives it its identity (Gelman, 2003). So when a bee is made to look like a spider, people can go beyond what it looks like and judge that it still has what makes it a bee—its essence. But what do people treat as the essence?

According to standard psychological essentialism, people represent essences as hidden underlying properties whose specific nature may be unknown but is assumed to be discoverable by science. For instance, people might elaborate the placeholder for the essence of a bee in terms of its DNA(see, e.g., Gelman, 2003; Keil, 1989). According to teleological essentialism, people represent essences in terms of purposes and so what something is for is treated as its essence. The essence of a bee is its purpose—making honey and pollinating flowers (see, e.g., Rose & Nichols, 2019).

Standard psychological essentialism and teleological essentialism diverge on a number of issues. Here we focus on one: the way in which manipulations of a thing's purpose affect categorization. According to standard psychological essentialism, essences are believed to cause category features. So if changing what something is for affects categorization judgments, then it must do so because people are reasoning diagnostically from a change in what something is for to a change in underlying causes. And it is those changes in the underlying causes that determine people's categorization judgments (Neufeld, 2021). In contrast, teleological essentialism maintains that manipulating a thing's purpose directly influences how it's categorized, even when a diagnostic inference is not licensed.

We begin by setting out standard psychological and teleological essentialism before turning to the "diagnosticity objection" against teleological essentialism. We then present evidence that the role of teleology in categorization can't be explained by diagnostic inference alone.

Psychological essentialism

Psychological essentialism—the view that humans represent some categories as having a true nature which is assumed to be shared among category members and responsible for similarities among category members—is highly influential (e.g., Gelman, 2003; Gelman & Wellman, 1991; Keil, 1989; Machery et al., 2023; Martin & Heiphetz, 2021; Medin & Ortony, 1989; Rhodes, Leslie, & Tworek, 2012). Psychological essentialism does not, on its own, make theoretical commitments on a range of issues. For example, it doesn't make a commitment on questions concerning the content of essences—are essences associated with scientific information or some other kind of information? Nor does it make a commitment on the structure of concepts—do essentialist concepts represent essences and observable

features in terms of causal models or some other kind of structure?[1] Finally, it does not make a commitment on what kinds are believed to have essences—are artifacts (e.g., clocks), biological kinds (e.g., bees), and non-living natural kinds (e.g., water) all treated as having essences? There are many possible versions of psychological essentialism, depending on how one answers these questions. However, one version has dominated the literature: standard psychological essentialism.

Standard psychological essentialism

Standard psychological essentialism has at least two central commitments. One is causal essentialism. The other is scientific essentialism.

Causal essentialism concerns conceptual structure. Essences are represented as underlying common causes that generate observable features of category members (Gelman, 2003; Medin & Ortony, 1989; Neufeld, 2022; Rehder, 2007, 2017). So, for instance, if we treat bees as having an essence, then that essence is viewed as generating the typical features of bees, such as stripes, wings, and stingers.

Scientific essentialism maintains that science uncovers the properties that make things what they are (e.g., Kripke, 1980; Putnam, 1975). While laypeople often don't know what these features are, they treat causal essences as placeholders that will eventually be filled out by scientific knowledge (Gelman, 2003; Keil, 1989). For instance, people might elaborate the placeholder for the essence of a bee with scientific information, thinking that its essence is its DNA (see, e.g., Gelman, 2003, pp. 138–139 and Neufeld, 2021, p. 5).

These two parts of standard psychological essentialism—causal essentialism and scientific essentialism—fit nicely together. For biological kinds at least, the scientific essence does cause the observable features. For instance, bee DNA causes bees to have stripes, wings, and stingers. Yet, while standard psychological essentialism has been the dominant view of essentialism for well over thirty years, recent work suggests that people sometimes fill in the placeholder of an essence with teleological information.

Teleological essentialism

Teleological essentialism is the view that people represent essences, at least in part, in terms of what they believe something is for (Rose & Nichols, 2019, 2020). In other words, what something is for—its purpose—is treated as its essence. On this view, placeholders are elaborated in terms of teleological information. With respect to bees, for instance, people tend to think that their purpose is making honey and pollinating flowers (Rose & Nichols, 2019). So if a bee is changed so that it no longer makes honey and pollinates flowers, but only spins webs to catch and kill insects, people think it no longer has the essence of bees but instead has the essence of spiders. Consequently, people are more likely to categorize it as a spider (Rose & Nichols, 2019).

Research has shown that across classic tests of essentialist thinking—from radical transformation, to switched-at-birth and offspring tasks—what something is for plays a role in determining what it is (see also, e.g., Rose, 2015; Rose & Nichols, 2020; Rose, Schaffer,

[1] We follow Neufeld (2021) in using "essentialist concepts" to refer to concepts that encode beliefs that some categories have essences. In doing so, we assume that not all concepts, for example "blue things", encode beliefs about essences.

& Tobia, 2020). For instance, if a bee has its insides completely replaced with the insides of a spider and now has the purpose of spiders, it is categorized as a spider. But if it has spider insides and the purpose of a bee, it is categorized as a bee (Rose, Jaramillo, Nichols, & Horne, 2022; Rose & Nichols, 2019). Moreover, if baby bees are placed in a cage full of spiders, and end up spinning webs when they mature, people judge that they are spiders (Rose & Nichols, 2019). And if a bee is transformed to look like a spider but preserves the purpose of a bee, adults predict that its offspring will be bees (Rose & Nichols, 2019). Together, these findings suggest that placeholders for essences are sometimes elaborated in terms of purposes. However, one key objection to the findings that support teleological essentialism is that they can be explained by familiar principles of diagnostic reasoning (Joo & Yousif, 2022; Neufeld, 2021).

The diagnosticity objection

The "diagnosticity objection" to the role of purposes in category membership judgments suggests that essence placeholders are not elaborated with teleology. Diagnostic reasoning involves making inferences from observed effects to their unobserved causes. If, as standard psychological essentialism proposes, some concepts are structured around beliefs about causal essences, then the influence of teleology on categorization can be explained in a way that remains compatible with scientific essentialism. On this view, a scientific essence is understood as the causal source of a category's observable properties. Observing those properties therefore licenses inferences about their hidden cause—namely, the essence itself. Consequently, if a bee is altered to resemble a spider but continues to make honey and pollinate flowers, participants may infer that, because it retains the purpose of bees, it must also retain whatever underlies that purpose: the relevant scientific essence, in this case, bee DNA.

If the diagnosticity objection were correct, it would imply that teleological information plays little or no role in how people represent essences. However, as Toorman (2023) argues, it is not clear that the diagnosticity objection succeeds. Classic transformation studies already involve properties that could, in principle, be treated as diagnostic evidence for changes in underlying essences. In Keil's original studies, for example, a raccoon is altered to look like a skunk. Younger children tend to judge that it has become a skunk, whereas older children tend to judge that it remains a raccoon (Keil, 1989). Standard psychological essentialism does not usually explain this developmental shift by appealing to diagnostic reasoning. That is, younger children are not typically said to infer from the animal's new appearance that its underlying scientific essence has changed, nor are older children said to stop treating appearance as evidence that an essence has changed. Rather, the transformation paradigm is taken to reveal which properties children and adults treat as relevant to category membership. The same point applies to cases involving teleology. If an organism's purpose is changed and participants judge that its category membership changes, that response need not show that they are merely using purposes as evidence that an underlying scientific essence has changed. It may instead show that they regard purpose itself as partly constitutive of what the thing is. Thus, unless there is some independent reason to treat transformation to a thing's purpose differently from perceptual transformations, the diagnosticity objection risks being ad hoc.

Showing that the diagnosticity objection risks being ad hoc, however, does not

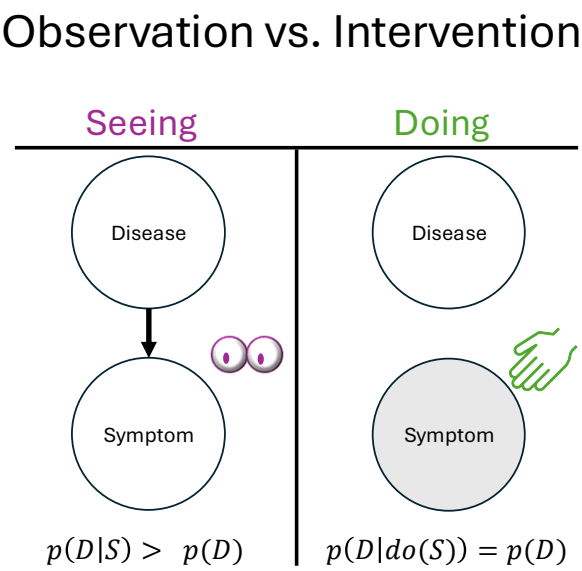

Figure 1

Diagnostic reasoning in situations involving observation and intervention. Diagnostic inference is licensed when an effect is observed. If a disease D (e.g., a cold) is present and causes some symptom S (e.g., fever), then the probability that a disease is present conditional on observing the symptom is greater than the probability of the disease without the symptom. But when the symptom is intervened on (e.g., one heats or cools one's forehead), it no longer provides evidence that the disease is present, as the intervention breaks the usual link between cause and effect.

establish that participants are not in fact reasoning diagnostically. We therefore test the diagnosticity objection empirically by capitalizing on a general principle of causal inference: whereas an *observed* effect provides evidence about its usual cause, a *produced* effect by an external intervention does not. If purpose continues to affect categorization when it is changed by an intervention while its usual underlying cause is held fixed, then diagnostic reasoning alone cannot explain the effect.

Diagnosticity and interventions

Our goal is to experimentally address the diagnosticity objection by presenting evidence that diagnostic reasoning alone cannot explain the influence of teleological considerations on categorization judgments. To do so, our strategy is to make use of causal interventions, since they undercut diagnostic inference, and explore whether teleology nonetheless affects categorization.

When reasoning about causes, different inferences are licensed depending on whether a variable was observed or intervened on (see Figure 1). For example, if a symptom is observed then it supports a diagnostic inference about its cause. If someone has a fever, for

instance, this licenses the inference to an underlying cause, in this case, that they have a cold.

Diagnostic reasoning relies on observations. Here, changes in effects imply changes in underlying causes. If a patient went from not having a fever to having a fever, then that would imply that the underlying cause has changed. If the patient had a cold, for instance, and this was what was causing their fever, then not having a fever would imply that they no longer had a cold.

However, diagnostic reasoning is undercut in cases of intervention. That is, diagnostic inference from observed effects to causes is not licensed when these effects were brought about through an external causal intervention (e.g., Sloman, 2005; Sloman & Lagnado, 2015; Waldmann & Hagmayer, 2005). If, for instance, someone heated their forehead, then we shouldn't infer that they have a cold. Interventions on a variable in a causal network break the incoming connections from its usual causes: the presence (or absence) of the symptom now doesn't provide any information anymore about how likely it is that the cause is present.

Interventions on effects break the links from their usual causes. As such, they undercut diagnostic inference. Transformation tasks involve interventions, and these tasks provide some of the best evidence of essentialist thinking. So, if differences in teleology affect categorization in transformation tasks even when it is clear that the purpose is brought about through an external intervention that leaves the usual causes intact, then this effect cannot be explained by diagnostic reasoning. We would then have evidence that in a central test of essentialist thinking, teleology matters in its own right.

We conducted two pre-registered experiments that use transformation tasks involving interventions to determine whether the effect of teleology on categorization judgments is explained by diagnostic reasoning alone (see Figure 2 for an overview). All analyses, data, materials, and links to pre-registrations are available here:

`https://anonymous.4open.science/r/teleology_beyond_diagnosticity_anonymous-1302`.

Experiment 1: Spiders with bee DNA

Experiment 1 addresses the diagnosticity objection by revisiting the central experiment that the objection was raised against. In this experiment, a bee is transformed to look like a spider and either makes honey or spins webs (Rose & Nichols, 2019). We use this experiment here but introduce one crucial addition: we add that when changing the bee to look like a spider, its DNA is held fixed, such that after the change, the creature still has bee DNA. If participants categorize a creature with the purpose of a spider and bee DNA as a spider, this couldn't be explained by diagnostic reasoning.

Methods

Participants

We recruited 100 participants (*age*: M = 36, SD = 14; *gender*: 48 female, 50 male, 2 non-binary; *race*: 1 American Indian/Alaskan Native, 18 Asian, 8 Black, 67 White, 6 no response/other) through Prolific. Participants received compensation at a rate of $12 an hour.

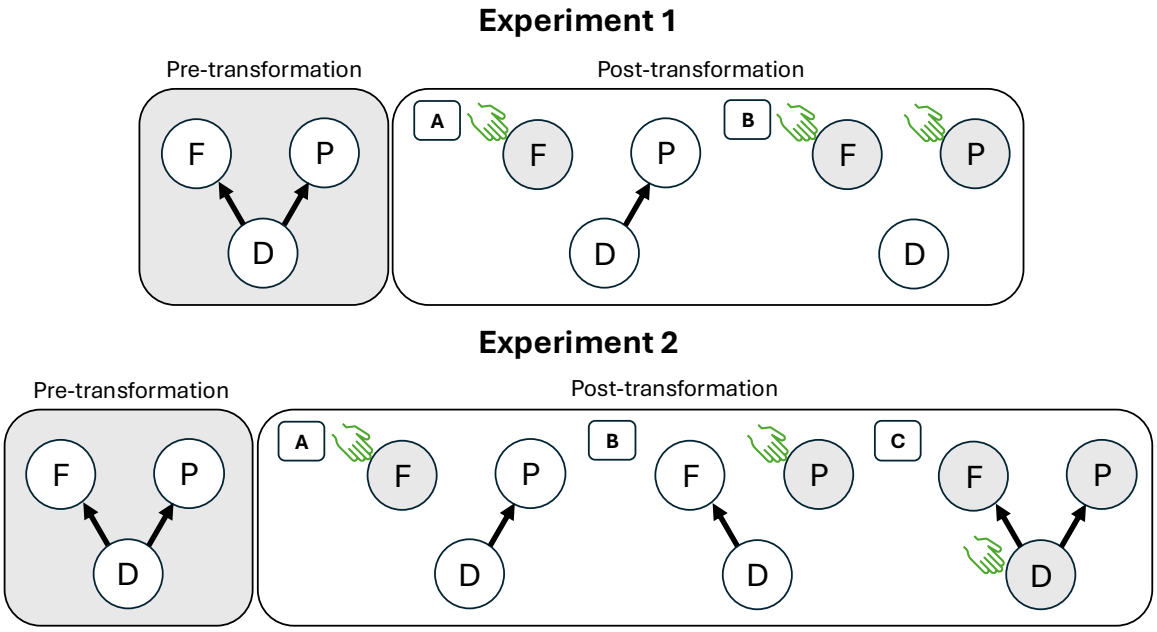

Figure 2

Overview of experiments. In Experiment 1, the pre-transformation animal is intervened on so that it (A) changes only its features (F) or (B) its features and purpose (P). The pre-transformation animal never has its DNA (D) changed. In Experiment 2, the pre-transformation animal is intervened on so that it either (A) changes only its features (F), (B) changes only its purpose (P), or (C) changes its DNA (D), which in turn changes both its features (F) and purpose (P).

Materials

Participants were told that some scientists make a bee look like a spider. In doing so, they do not alter the DNA of the original bee. The materials are shown in Figure 3.

Design and procedure

The experiment was programmed in jsPsych (de Leeuw, Gilbert, & Luchterhandt, 2023). Participants were randomly assigned to one of two between-subjects conditions: Bee Purpose Changed or Bee Purpose Preserved. In both conditions, participants read a vignette in which a bee was transformed to look like a spider. The conditions differed only in what happened to the animal's purpose after the transformation. In the Bee Purpose Changed condition, the transformed animal came to have a spider purpose, namely spinning webs. In the Bee Purpose Preserved condition, the transformed animal retained the original bee purpose, namely making honey.

Participants then answered two comprehension questions on a separate page.[2] After

[2]The two comprehension questions were, "Suzy and Andy performed a special operation on a bee.", and "The thing after the special operation only spins webs to catch insects and eat them.". Participants responded "yes" or "no". After responding to both comprehension questions, they proceeded to the test

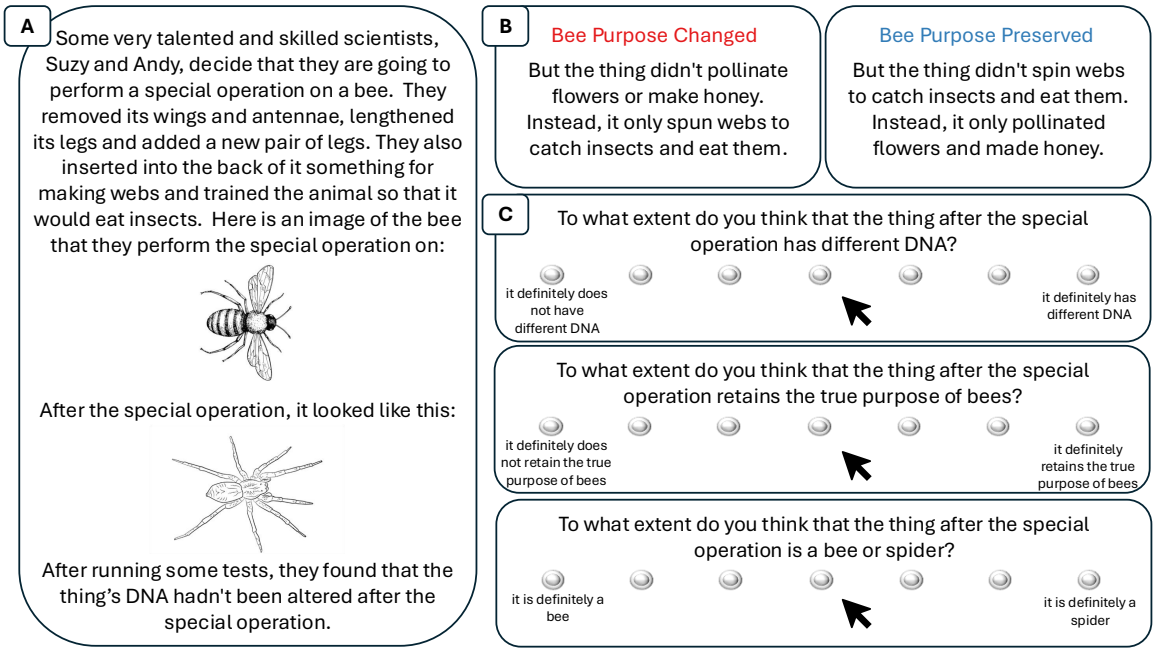

Figure 3

Experiment 1 overview. (A) Participants were told that some scientists make a bee look like a spider without altering its DNA. (B) They then learned that after the procedure the original bee's purpose was changed to that of a spider ("Bee Purpose Changed") or that it had the purpose of bees ("Bee Purpose Preserved"). (C) Participants were then asked whether its DNA was changed, whether it retained the purpose of bees, and whether it was a bee or spider.

these comprehension checks, participants completed the key categorization measure, indicating whether the transformed animal was a bee or a spider. They then answered two follow-up questions assessing whether the animal's DNA had changed and whether it retained the purpose of bees; the order of these two follow-up questions was randomized. All responses were made on 7-point scales (see Figure 3).

Results

Our pre-registered hypothesis was that changes in the creature's purpose would affect categorization judgments even when it is stipulated that the creature's DNA has not been changed. We first dummy coded our condition variable so that 0 = purpose preserved and 1 = purpose changed. We then fit a Bayesian mixed effects model with purpose as a predictor of category judgments, using the default priors from brms (Buerkner, 2017). In reporting our results throughout, we will refer to a statistical result of interest as "credible" when the 95% credible interval excludes 0.

The results are shown in Figure 4. As predicted, we found a credible effect of purpose on participants' judgments. The posterior estimate for purpose was positive, indicating that

trial.

in the purpose changed condition, participants were inclined to think the thing changed categories (i.e., the original bee is now a spider), and the 95% credible interval excluded 0, $M = 2.25,\ 95\%$ credible interval $= [1.52, 2.98]$.

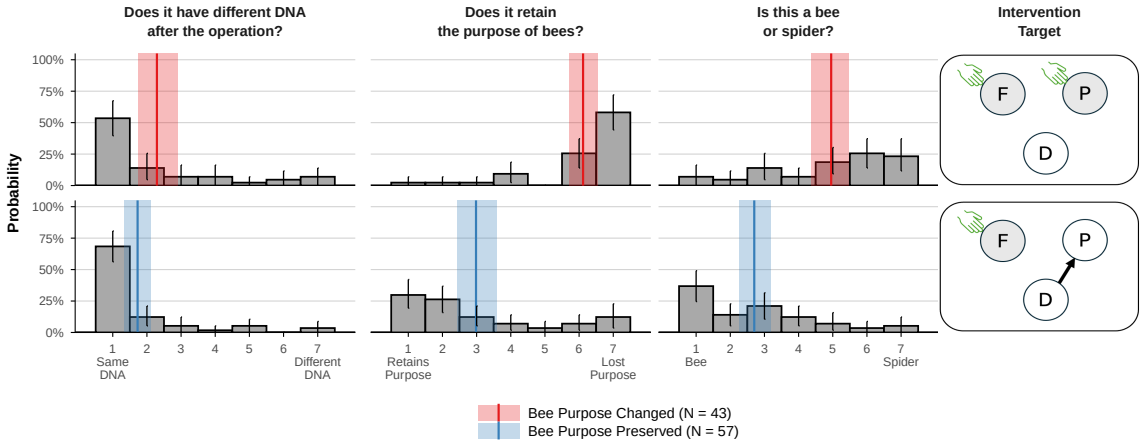

Figure 4

Response distributions for the DNA, purpose, and categorization questions based on whether the original bee preserved or changed its purpose after the procedure, along with the causal model—with D (DNA), P (Purpose), and F (Feature)—and interventions on variables for each condition. Error bars are 95% bootstrapped confidence intervals. Vertical lines are means with 95% bootstrapped confidence intervals as ribbons.

Our manipulations were also effective. To test this, we fit two additional models, one with condition as a predictor of DNA and one with condition as a predictor of purpose. The 95% credible interval of the posterior for DNA did not exclude 0, $M = .57,\ 95\%$ credible interval $= [-0.12, 1.24]$, indicating that in both cases, people were similarly inclined to judge that the creature had bee DNA. The 95% credible interval of the posterior for purpose excluded 0, $M = 3.13,\ 95\%$ credible interval $= [2.37, 3.89]$, indicating that people were more inclined to think that the creature had the bee purpose when it still made honey and pollinated flowers after the procedure.

Discussion

We found that a bee that was changed to have the purpose of a spider—spinning webs—but still had bee DNA was categorized as a spider. Because DNA was explicitly held fixed, this effect cannot be explained by participants inferring from the change in purpose that the creature's DNA had changed. Nonetheless, the change in purpose affected categorization judgments.

However, participants could still have inferred that something else internal to the animal had changed. Even so, this doesn't support the diagnosticity objection as usually formulated. The objection is not merely that participants think some internal change occurred; it is that teleological information affects categorization only because it provides evidence for an underlying scientific essence, such as DNA. Still, the broader possibility remains that

participants interpreted the purpose change as evidence for some other unmeasured internal change.

Experiment 2 was designed to address this possibility. There, we made the relevant causal structure explicit, represented DNA as the common cause of both a creature's purpose and a non-functional feature, and tested whether interventions on purpose continued to have a privileged effect on categorization when DNA was held fixed.

Experiment 2: Interventions in an explicit common-cause structure

Experiment 2 explicitly represents a common-cause structure with DNA as the root cause and two effects of that cause. One effect involves a creature's purpose (e.g., making honey) while the other involves a superficial feature (e.g., having red spots). Having explicitly represented the causal structure, we will then explicitly represent interventions on variables in the model. If explicitly representing that the purpose and feature occupy the same position in a causal model—both are effects of a common cause—and have exactly the same number of inputs and outputs, we find that interventions on the purpose play a greater role in categorization than interventions on the feature, then that would continue to suggest that the effect of teleology on categorization can't be explained by diagnostic reasoning.

Methods

Participants

We recruited 60 participants (*age*: M = 24, SD = 4.9; *race*: 1 Asian, 5 Black, 51 White, 3 no response/other) through Prolific. Gender was not collected. Participants received compensation at a rate of $11 an hour.

Materials

The materials involved two novel creatures, a Xan and a Zerp (see Figure 5). For both Xans and Zerps, their DNA caused their purpose and feature. Xans were described as having red spots, the purpose of making honey, and as having XYZ DNA; Zerps were described as having blue spots, the purpose of spinning webs, and as having ABC DNA.

Procedure

The experiment was programmed using jsPsych (de Leeuw, 2015). Participants were first separately introduced to Xans and Zerps and told—as well as shown—that they have particular causal structures (see Figure 5).

After learning about the two creatures, participants were told that two talented and skilled scientists, Suzy and Andy, would be performing operations on the creatures. They were then given five comprehension questions to make sure they understood the differences between Xans and Zerps. If they missed one of the comprehension questions, they would return to the beginning of the experiment, read about Xans and Zerps again, and then complete the comprehension questions again. When all comprehension questions were answered correctly, participants then proceeded to the testing phase. Here they were given three test trials, in random order, that varied in what was intervened on.

In one trial, the purpose was intervened on; in another trial, the spots were intervened on. Lastly, we included, as a control, a trial where the DNA was intervened on which led to both its purpose and features being changed (see Figure 5).

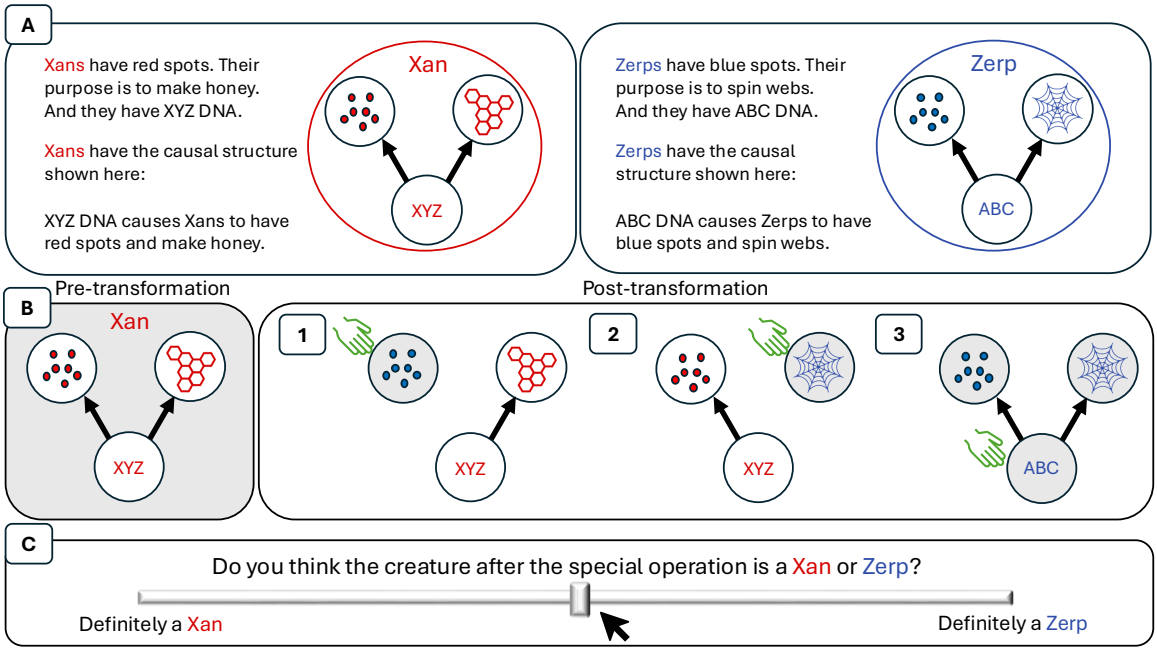

Figure 5

Experiment 2 overview. (A) Participants are introduced to Xans and Zerps. Xans are depicted and described as having XYZ DNA that causes them to have red spots and make honey. Zerps are depicted and described as having ABC DNA that causes them to have blue spots and spin webs. (B) They then proceed to test trials where interventions are made. Here the pre-transformation animal is a Xan. Participants see interventions on (1) its spots, (2) its purpose, and (3) its DNA. (C) In each trial, they indicate whether the post-transformation animal is a Xan or Zerp.

On each trial participants were asked whether the creature was a Xan or Zerp. All ratings were made on a slider that ranged from 0 to 100 (the numbers were not displayed) with anchors on each end that were either "definitely a Xan" or "definitely a Zerp".

Design

The order of the three trials was randomized. We also counterbalanced whether a Xan or a Zerp underwent transformation. When Zerps were transformed, participants were told that Suzy and Andy would be performing special operations on Zerps and the rating scale was anchored to the left with "definitely a Zerp". When Xans were transformed, participants were told that Suzy and Andy would be performing special operations on Xans and the rating scale was anchored to the left with "definitely a Xan". The slider handle only appeared when participants clicked on the scale.

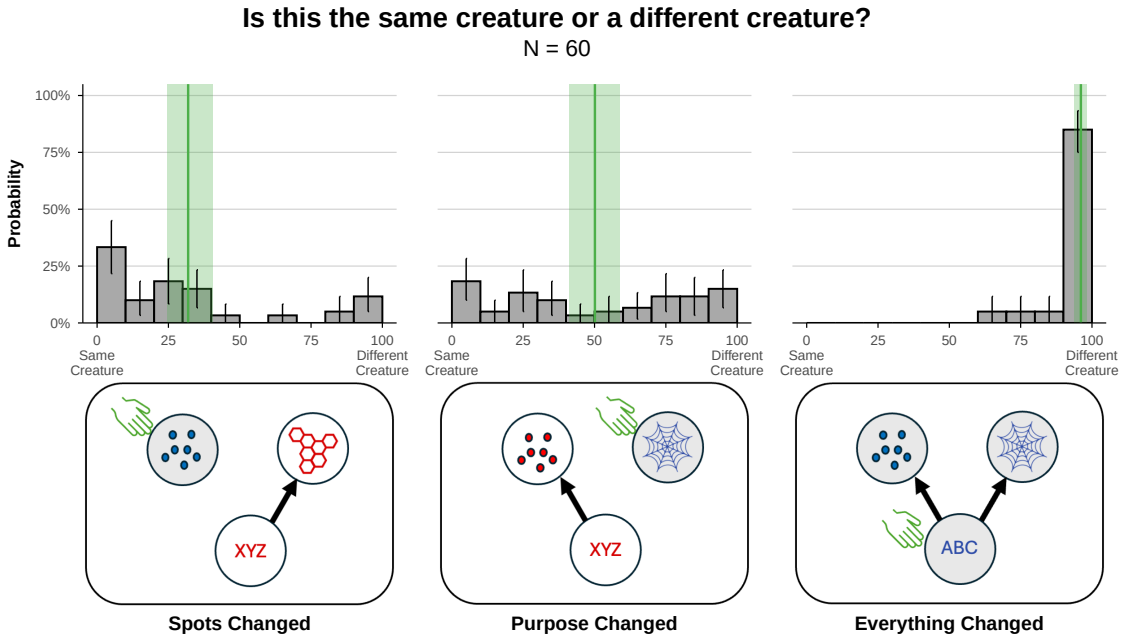

Figure 6

Response distributions for trials where only spots were changed, only the purpose was changed, or everything was changed. Error bars show 95% bootstrapped confidence intervals. The green vertical lines are means, and the ribbons show 95% bootstrapped confidence intervals.

Results

Our pre-registered hypothesis was that interventions on the creature's purpose would have a greater effect on categorization than interventions on the creature's features, even though it is made explicit that the DNA is held fixed. We also predicted that since our control condition involved intervening on the DNA and this led to changes in both the purpose and features—a total transformation—this would have the greatest effect on categorization judgments.

We fit a Bayesian regression model with intervention as a predictor of categorization judgments. We then tested, using emmeans, whether 95% of the posterior distribution of the difference between purpose and spots was greater than 0 (see Figure 6). We found that it was $M = 18.2$, 95% credible interval $= [10.4, 26.9]$. This indicates that interventions on purpose had a greater effect on categorization judgments than intervention on spots. We also found that 95% of the posterior distribution of the difference between DNA and spots, $M = 64.3$, 95% credible interval $= [56.5, 73.1]$, and DNA and purpose, $M = 46.0$, 95% credible interval $= [37.2, 53.7]$, were greater than 0. As expected, when a creature is completely changed, this has a greater effect on categorization.

Discussion

When explicitly representing a common-cause structure in which DNA jointly determined a creature's purpose and its spots (a superficial feature), interventions altering the creature's purpose exerted a greater influence on categorization than interventions altering its spots. This pattern is difficult to reconcile with a diagnostic reasoning account. Because both purpose and spots were represented as effects of the same underlying cause and occupied equivalent positions within the causal structure, differences in diagnostic inference cannot explain why purpose changes mattered more.

However, perhaps participants assumed that DNA was more strongly connected to purpose than to spots. If so, a change in purpose would provide stronger evidence for a change in DNA than a change in spots would (e.g. because there are other factors than DNA that can affect the spots but not the purpose). But purpose and spots were both represented as effects of DNA, and the relevant changes were explicitly described as interventions. Since interventions on effects undercut diagnostic inference to their usual causes, an externally induced change in purpose should not provide evidence that DNA changed simply because DNA normally causes purpose.

As expected, our control condition showed that altering the common cause produced the largest shift in categorization. In the model, changing the cause changes its downstream effects. Thus, interventions on DNA altered every variable in the structure, whereas interventions on purpose or spots altered only a single property.

The results of Experiment 1 showed that when DNA was explicitly held fixed, changing a creature's purpose influenced categorization judgments. Experiment 2 extends this finding by demonstrating that the privileged influence of purpose persists even when purpose and features are represented as effects of a shared underlying cause. Together, these findings indicate that teleological effects on categorization cannot be explained by participants inferring changes in DNA from changes in purpose.

General Discussion

A central challenge for teleological essentialism is to show that the influence of purpose on categorization is not merely diagnostic. Prior work suggests that people sometimes categorize things by what they are for. But this pattern admits an alternative explanation: perhaps people use purpose only as evidence for an underlying scientific essence, such as DNA. On this diagnosticity objection, when a creature's purpose changes, people infer that its underlying essence has changed, and it is this inferred change in essence—not purpose itself—that drives categorization.

We tested this objection by asking whether purpose still affects categorization when the relevant diagnostic inference is blocked. Across two experiments, it did. In Experiment 1, participants categorized a transformed bee as a spider when it acquired the purpose of spiders, even though its DNA was explicitly held fixed. In Experiment 2, we made the causal structure explicit: DNA caused both a creature's purpose and a non-functional feature. Even then, intervening on the creature's purpose had a greater effect on categorization than intervening on the feature, despite both being effects of the same underlying cause. Together, these findings suggest that purpose affects categorization in a way that cannot be explained by diagnostic reasoning alone.

Our experimental strategy turned on the fact that interventions on effects block the relevant diagnostic inference to their usual causes. If a creature is merely observed to spin webs, that may provide evidence that it has the underlying essence of spiders. But if the creature is made to spin webs by an external intervention, and its DNA is explicitly held fixed, then spinning webs no longer provides evidence that its DNA has changed. This is the crucial point for the diagnosticity objection: when a purpose is changed by intervention while its usual underlying cause is held fixed, a continued effect of purpose on categorization cannot be explained by an inference from purpose to underlying scientific essence (e.g., Sloman, 2005; Sloman & Lagnado, 2015; Waldmann & Hagmayer, 2005).

Experiment 1 used this strategy to revisit the central kind of case that motivated the diagnosticity objection. Participants learned that scientists transformed a bee so that it looked like a spider, while explicitly leaving its DNA unchanged. When the transformed creature retained the purpose of bees, participants tended to categorize it as a bee. But when it acquired the purpose of spiders, participants tended to categorize it as a spider. Because the creature's DNA was stipulated to remain bee DNA, this effect cannot be explained by participants inferring a change in DNA from a change in purpose. Nonetheless, changing the creature's purpose affected categorization.

Experiment 2 extended this result by making the causal structure explicit. Participants learned about novel creatures whose DNA caused both a purpose and a physical feature. This allowed us to compare interventions on two effects of the same underlying cause. If purpose influenced categorization only because it was diagnostic of DNA, then an intervention on purpose should not have a privileged effect once participants were shown that the purpose and feature occupied equivalent positions in the causal structure. Yet interventions on purpose shifted categorization judgments more than interventions on spots. Thus, even when purpose and a non-functional feature were both represented as effects of a common cause, purpose played a larger role in determining category membership (see also Lombrozo & Rehder, 2012).

The control condition in Experiment 2 also behaved as expected. When DNA itself was changed, participants showed the largest shift in categorization. This is unsurprising: in the explicit causal model, DNA was the common cause of both the creature's purpose and its feature. Intervening on DNA therefore changed the whole causal structure, whereas interventions on purpose or spots changed only one downstream property. This pattern suggests that participants were sensitive to the causal model. But crucially, among the downstream effects, purpose mattered more than a non-functional feature.

Together, these findings suggest that the effect of teleological considerations on categorization cannot be explained by diagnostic reasoning alone. This does not mean that diagnostic reasoning never plays a role in categorization. Nor does it mean that people never categorize biological kinds by appeal to underlying scientific causes. Indeed, our results are compatible with the idea that people sometimes use purposes diagnostically, and that they sometimes treat DNA or other hidden causes as central to category membership. The point is instead that diagnostic reasoning is not sufficient to explain the role of purpose in these cases. Even when the relevant diagnostic inference is blocked, purpose continues to affect categorization.

These findings therefore support teleological essentialism. They suggest that people do not merely use what something is for as evidence for some deeper, scientifically construed

essence. Rather, purpose can itself play a role in how people determine what something is. At least in some central tests of essentialist categorization, what something is for matters in its own right.

Conclusion

Teleological essentialism is the view that people treat what something is for—its purpose—as part of its essence. Recent work supporting teleological essentialism has been argued to be explained by people making diagnostic inferences from changes in purpose to changes in underlying causes of that purpose, namely, DNA. Here, we showed that effects of teleology on categorization can't be explained by diagnostic reasoning alone. This suggests that the role of teleology in categorization isn't straightforwardly explained in a way that is entirely compatible with standard psychological essentialism. Instead, these results provide support for teleological essentialism.

References

Buerkner, 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

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. doi: 10.21105/joss.05351

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213–244. doi: 10.1016/0010-0277(91)90007-Q

Joo, S., & 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.

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

Machery, E., Olivola, C. Y., Cheon, H., Kurniawan, I. T., Mauro, C., Struchiner, N., & Susianto, H. (2023). Is identity essentialism a fundamental feature of human cognition? *Cognitive Science*, *47*(5), e13292. doi: 10.1111/cogs.13292

Martin, J. W., & Heiphetz, L. (2021). "internally wicked": Investigating how and why essentialism influences punitiveness and moral condemnation. *Cognitive Science*, *45*(6), e12991. doi: 10.1111/cogs.12991

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: 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 *Mind, language and reality: Philosophical papers, volume 2* (pp. 215–271). Cambridge: Cambridge University Press.

Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195176803.003.0013

Rehder, B. (2017). Concepts as causal models: Categorization. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 347–376). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199399550.013.39

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. doi:

10.1073/pnas.1208951109

Rose, D. (2015). Persistence through function preservation. *Synthese*, *192*(1), 97–146. doi: 10.1007/s11229-014-0555-6

Rose, D., Jaramillo, S., Nichols, S., & Horne, Z. (2022). Teleological essentialism across development. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44). Retrieved from `https://escholarship.org/uc/item/0z45p9zs`

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

Rose, D., Schaffer, J., & Tobia, K. (2020). Folk teleology drives persistence judgments. *Synthese*, *197*(12), 5491–5509. doi: 10.1007/s11229-018-01974-0

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.

Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology*, *66*, 223–247. doi: 10.1146/annurev-psych-010814-015135

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

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 216–227. doi: 10.1037/0278-7393.31.2.216