

Contextual utility affects the perceived quality of explanations

Nadya Vasilyeva¹ · Daniel Wilkenfeld² · Tania Lombrozo¹

Published online: 6 April 2017
© Psychonomic Society, Inc. 2017

Abstract Are explanations of different kinds (formal, mechanistic, teleological) judged differently depending on their contextual utility, defined as the extent to which they support the kinds of inferences required for a given task? We report three studies demonstrating that the perceived “goodness” of an explanation depends on the evaluator’s current task: Explanations receive a relative boost when they support task-relevant inferences, even when all three explanation types are warranted. For example, mechanistic explanations receive higher ratings when participants anticipate making further inferences on the basis of proximate causes than when they anticipate making further inferences on the basis of category membership or functions. These findings shed light on the functions of explanation and support pragmatic and pluralist approaches to explanation.

Keywords Explanation · Inference · Inductive utility · Context · Pragmatic factors

Electronic supplementary material The online version of this article (doi:10.3758/s13423-017-1275-y) contains supplementary material, which is available to authorized users.

✉ Nadya Vasilyeva
vasilyeva@berkeley.edu

Daniel Wilkenfeld
daw171@pitt.edu

Tania Lombrozo
lombrozo@berkeley.edu

¹ Department of Psychology, University of California Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA

² Department of History and Philosophy of Science, University of Pittsburgh, 1017 Cathedral of Learning, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA

Do people evaluate the quality of explanations differently depending on how well the explanations suite their needs in a given context? Suppose, for instance, that Ana and Bob are both interested in marsupials. Ana is studying marsupials to diagnose their ailments; Bob is interested in understanding their adaptations. When it comes to explaining why kangaroos have large tails, will Ana find mechanistic explanations (for instance, in terms of development or genes) more compelling than Bob? Will Bob find teleological explanations (for instance, that appeal to balance) more compelling than Ana?

Research increasingly supports the idea that (many) representations and judgments are sensitive to contextual factors, including an individual’s goals and the task at hand (e.g., Aarts & Elliot, 2012; Barsalou, 1983; Markman & Ross, 2003). This raises the possibility that judgments concerning the quality of explanations are similarly flexible. Moreover, some accounts of explanation can naturally accommodate forms of context sensitivity. Lombrozo and Carey (2006), for example, suggest that one function of explanation is to support future reasoning and behavior by highlighting generalizable or “exportable” relationships (see also Craik, 1943; Heider, 1958). Given that information is differentially useful in different contexts, one might expect judgments of explanation quality to reflect contextual utility: Explanations should be perceived as better to the extent they contain information that’s inductively useful given one’s current or expected context (see also Leake, 1995, for a relevant discussion).

Within philosophy, so-called pragmatic accounts of explanation also allow for the possibility of substantial context sensitivity. For example, van Fraassen (1980) proposes that context shapes the contrast class—that is, the set of possible alternatives to the target observation that the explanation needs to account for—as well as the relevance relationship between the explanation and what it explains (that is, the relationship that makes the answer explanatory with respect to a particular

question in a particular context; see also Gorovitz, 1965; Hilton, 1990; Hilton & Erb, 1996; Lipton, 1990, 1993 for theories of explanatory relevance). Asking whether kangaroos have *long* tails (as opposed to *short* tails) involves a different contrast class from asking why they have long *tails* (as opposed to, say, long *front paws*), whereas providing a mechanistic explanation arguably involves a different relevance relationship from a teleological explanation. Such proposals raise the possibility that different contexts call for different *kinds* of explanations to account for one and the same observation. More concretely: Ana might be right, given her context, to favor a mechanistic explanation for the kangaroo's tail, and Bob might be right, given his context, to favor a teleological explanation.

In the current studies we test the hypothesis that judgments of explanation quality are sensitive to contextual utility. More specifically, we test the prediction that explanations of a given kind (e.g., mechanistic vs. teleological) will receive a relative boost when they highlight explanatory relationships that are likely to support the evaluator's inductive aims in the context of a given task. In the remainder of the introduction, we clarify our notion of "contextual utility," briefly review past work on related questions, and provide an overview of the three experiments we go on to report.

Explanations in context

Explanation generation and evaluation are affected by many factors, including the explainer or evaluator's beliefs (e.g., Hilton, 1990; Pennington & Hastie, 1993), the intended recipient of the explanation (e.g., Vlach & Noll, 2016), and social/motivational considerations, such as "saving face" or persuasion (see Patterson, Operskalski, & Barbey, 2015, for review). Explanations are also generated and evaluated in contexts that involve various (potentially inconsistent) goals operating at multiple scales. In explaining why kangaroos have long tails, Ana might want to diagnose a specific medical condition, understand marsupial physiology, do well on a veterinary exam, and impress her instructor all at once. While these forms of "context sensitivity" are of interest in their own right, the focus of the present research is on the extent to which an explanation of a particular kind is privileged in virtue of highlighting a generalization that is relevant to the recipient's immediate task. In other words, we examine whether the perceived value of explanations is proportional to the degree of guidance they provide for inferences anticipated in the current context. This is what we refer to by contextual utility. To isolate this facet of explanation, it is important to keep other factors fixed, including the question being asked, the source of the explanation, and the background knowledge of the participant evaluating a given explanation.

Prior work provides evidence that the evaluation and generation of explanations is indeed sensitive to context, but

context has been varied alongside participants' explanatory task or background beliefs. For example, shifts in contrast class have been shown to influence both the generation and the evaluation of causal explanations (McGill, 1989; Hilton & Erb, 1996), but in such cases the explanations are effectively answering different questions. Chin-Parker and Bradner (2010) found that the frequency with which participants generated mechanistic and teleological explanations for an event sequence was influenced by changing background conditions, but this manipulation also changed background knowledge. Finally, Hale and Barsalou (1995) had participants complete a task with an initial system-learning phase followed by a trouble-shooting phase, and found that the types of explanation that participants generated varied across phases. However, phase was confounded with several factors, including task order, changes in background knowledge, and task instructions (think aloud vs. explanation). It thus remains an open question whether contextual utility—as we have defined it—affects the perceived quality of explanations.

If anything, research to date suggests that when all of these factors are held constant, explanatory preferences are quite stable (e.g., Kelemen, Rottman, & Seston, 2013; Lombrozo, 2007). Moreover, mainstream accounts of explanation from philosophy have typically set pragmatic and contextual considerations to the side, instead focusing on a specification of formal relationships or features that are constitutive of explanations, such as deductive arguments of a particular form (Hempel & Oppenheim, 1948) or causal processes that generate an effect (Salmon, 1984). On these views, pragmatic factors have a limited influence, perhaps in what one chooses to explain or in the level at which an explanation is pitched.

Here we investigate whether people evaluate an explanation differently depending on its contextual utility, that is, the degree of *guidance* the explanation is expected to provide for the kinds of inferences that a person anticipates making in the context of a given task. We aim to provide a direct test of this effect by keeping background knowledge and what is being explained fixed across evaluators' tasks. We expect any effects, if found, to be relatively small, as they should operate on top of relatively stable preferences determined by the features that are held constant across our experimental manipulations.

We report three experiments in which we ask people to evaluate explanations of different kinds: formal (which appeal to category membership; see Prasada & Dillingham, 2009a), mechanistic (which appeal to proximate causes), and teleological (which appeal to goals or functions). Importantly, we experimentally manipulate the contextual utility of each kind of explanation by varying whether the relationships that underwrite each type of explanation—that is, between a property and category membership (formal), its proximate causes (mechanistic), or its function (teleological)—are more or less useful in light of participants' task. If judgments of explanation quality are sensitive to the contextual utility of the

generalization that underwrites a given explanation, then formal, mechanistic and teleological explanations should receive higher ratings in the context of tasks involving generalizations along corresponding dimensions. For example, if a task

requires participants to predict the presence of a given feature on the basis of its function (as opposed to its category membership or a proximate cause), this should make participants value generalizations that relate the feature to the function (e.g., “long

Table 1 Sample instructions, explanations, and task-reinforcers used in Experiment 1. Task was manipulated between subjects; explanation type and domain were manipulated within subjects (with one explanation presented for evaluation on each trial). Text referring to artifacts is provided in square brackets

1. Generalization instructions (Task manipulation)

(After evaluating each explanation) you will make a *prediction* about a new object or organism:

Category-based task:

The new organism [object] will *either be of the same species [kind] as each original that you read about or of a different species [kind]*.

Cause-based task:

The new organism [object] will *either have the same internal characteristics [parts] and processes as each original that you read about or have different internal characteristics [parts] and processes*.

Function-based task:

The new organism [object] will *either have the same needs [purpose] as each original that you read about or have different needs [purpose]*.

Based on this information you'll have to guess whether the object or organism has the same properties as the original or different properties.

2. Description



Glentas are microorganisms in the ocean. Their motion is controlled by a set of light-seeking photoreceptors, which makes them rise towards the ocean's surface during the day. Spending some time at the ocean's surface helps them replenish their oxygen reserves.

3. Explanation evaluation

Below is a picture of one particular specimen, ID-Zd89u0002, from a research facility. Why does this specimen *rise to the ocean's surface during the day*?



Formal explanation:

Because it's a glenta, and glentas rise to the ocean's surface during the day.

Mechanistic explanation:

Because its motion is controlled by a set of light-seeking photoreceptors, which makes it rise to the ocean's surface during the day.

Teleological explanation:

Because rising to the ocean's surface during the day helps it replenish oxygen reserves.

Circular explanation:

Because some things can rise to the ocean's surface.

Very bad explanation (1) – Very good explanation (9)

4. Task reinforcement

Behind this box there is a microorganism. Click [HERE](#) to find out...



Category-based task:

...if it's a glenta.
(Text appears on click)
Yes, it's a glenta *or*
No, it's NOT a glenta.

Cause-based task:

...if its motion is controlled by a set of light-seeking photoreceptors.
(Text appears on click) Yes, its motion is controlled by a set of light-seeking photoreceptors *or* No, it its motion is NOT controlled by a set of light-seeking photoreceptors.

Function-based task:

...if it needs to replenish oxygen reserves.
(Text appears on click) Yes, it needs to replenish oxygen reserves *or* No, it does NOT need to replenish oxygen reserves.

Do you think it rises to the ocean's surface during the day?

Definitely no (1) - Definitely yes (9)

tails support balance”). Because this is a generalization that underwrites a teleological explanation (“kangaroos have long tails because they improve balance”), the perceived quality of teleological explanations in this context should be boosted relative to their perceived quality in contexts involving category-based or cause-based generalizations.

Experiment 1

Participants learned about novel artifacts and biological kinds with target features that supported multiple explanations. For instance, participants read about a microorganism with a property (rises to the ocean's surface) supporting a formal explanation (because it is a *glenta*), a mechanistic explanation (because it has special photosensitive receptors), and a teleological explanation (because doing so helps it replenish oxygen reserves). These explanations were evaluated in the context of a generalization task that required participants to predict the presence of the target feature in a new item based on information about either its category membership (category-based task), its proximate causal structure (cause-based task), or its functions (function-based task). As additional reference points, we included circular explanations and a baseline condition in which participants evaluated explanations in the absence of any additional task. We predicted that ratings of explanation “goodness” would be affected by the type of generalization that the specified task involved, with a boost for explanations congruent with that task.

To isolate the effects of contextual utility (as opposed to background knowledge), all participants received the same information about the target phenomena. Thus, all explanations (formal, mechanistic, and teleological) drew from the same pool of stated facts but pointed out different regularities that would support different generalizations (i.e., inferences based on shared category, cause, or function).

Method

Participants Four-hundred-and-twelve participants were recruited on Amazon Mechanical Turk in exchange for \$1.65; an additional 95 participants were excluded for failing a memory check that consisted of classifying descriptions of living things and artifacts as seen vs. unseen (12 descriptions in Experiments 1 and 3, allowing for up to two errors; 36 descriptions in Experiment 2, allowing for up to six errors). In all experiments, participation was restricted to workers with an IP address within the United States and with an approval rating of 95% or higher from at least 50 previous tasks on Mechanical Turk.

Materials, design, and procedure Participants were presented with descriptions of 16 fictional living things and artifacts, each described with a label and three features organized into a

causal chain (see Table 1 for an example, and [Supplementary Materials](#) for the full list of stimuli). For each entity, participants evaluated one of four explanations for the middle feature in the causal chain (formal, mechanistic, teleological, or circular) using a 9-point scale anchored at *very bad explanation* (1) and *very good explanation* (9). All explanations cited information familiar from the item description. During training, participants were specifically instructed to rate explanation *goodness* rather than truth (see [Supplementary Materials](#) for details). Each participant evaluated four explanations of each type, with item-explanation pairings counterbalanced across participants.

Crucially, participants rated explanations in either a baseline condition, which did not involve an additional task, or in a generalization condition that specified one of three additional tasks: category-based, cause-based, or function-based generalization. In each generalization condition, participants were informed that after evaluating explanations (as illustrated with two training trials), they would be making predictions about new objects and organisms, where the predictions would be based on known category membership, cause features, or function features. This served as the manipulation of contextual utility, and it was reinforced after each explanation evaluation by having participants perform an inference of the promised type. Specifically, participants were given information about an entity behind a black box and had to rate how likely it was that the target feature of the original item generalized to the occluded item (see Table 1 for examples). The information provided varied across tasks: participants were told whether the occluded entity belonged to the same category as the original (category-based), shared the same cause feature (cause-based), or shared the same function (function-based). The main purpose of this step was to maintain participants' focus on the task. Ratings were therefore not analyzed and are not reported.

Participants completed 16 trials, each consisting of an explanation evaluation (in which they rated the quality of an explanation), and for participants in one of the three generalization conditions, a subsequent prediction to reinforce the specified task. Because domain was not a variable of central theoretical interest, and because it did not interact with the effect of task in Experiments 1 or 2, we collapsed across this variable for analyses.

Results and discussion

Explanation ratings were analyzed in an ANOVA with explanation type as a within-subjects factor and task as a between-subjects factor. This revealed significant main effects of both explanation type, $F(3, 1224) = 1365.60, p < .001, \eta_p^2 = .770$, and task, $F(3, 408) = 6.81, p < .001, \eta_p^2 = .048$. Overall, participants preferred mechanistic and teleological explanations over formal explanations, all of which were preferred

Table 2 Mean explanation goodness ratings as a function of explanation type and task in Experiments 1, 2 and 3 (SDs in brackets)

Task (Exp.1, 2) or Prime (Exp. 3)	Explanation type				
	Formal	Mechanistic	Teleological	Circular	All
Experiment 1					
Category-based	4.03 (1.95)	7.46 (1.41)	7.27 (1.52)	2.04 (1.31)	5.19 (0.96)
Cause-based	3.64 (1.74)	7.44 (1.41)	6.57 (1.73)	1.85 (1.05)	4.88 (0.80)
Function-based	3.52 (1.91)	6.64 (1.75)	7.49 (1.40)	1.97 (1.27)	4.91 (0.94)
Baseline	3.04 (1.47)	6.76 (1.55)	7.22 (1.43)	1.63 (0.98)	4.66 (0.73)
All task conditions	3.56 (2.43)	7.07 (1.58)	7.15 (1.55)	1.88 (1.17)	
Experiment 2					
Categorize	3.88 (2.43)	7.38 (1.69)	7.23 (1.70)	2.08 (1.62)	5.14 (1.07)
Identify causal origin	3.8 (2.39)	7.29 (1.88)	6.90 (2.07)	2.13 (1.51)	5.03 (1.13)
Identify function	3.43 (2.04)	6.19 (2.33)	7.52 (1.78)	1.99 (1.44)	4.78 (1.13)
Baseline	3.42 (2.05)	6.95 (2.04)	7.03 (1.99)	2.06 (1.27)	4.87 (1.03)
All task conditions	3.63 (2.23)	6.95 (2.08)	7.17 (1.90)	2.07 (1.46)	
Experiment 3					
Inductively useful causes	5.03 (2.36)	7.97 (1.40)	5.95 (2.30)	2.09 (1.60)	5.26 (1.00)
Inductively useful functions	4.06 (2.02)	6.78 (1.96)	6.13 (1.83)	2.22 (1.46)	4.80 (1.10)
Salient causes	3.94 (2.27)	6.57 (2.16)	6.46 (2.28)	1.96 (1.45)	4.73 (1.10)
Salient functions	3.89 (2.15)	6.70 (2.07)	6.52 (2.07)	2.18 (1.61)	4.82 (1.19)
No prime (baseline)	3.72 (1.82)	6.81 (1.79)	7.17 (1.53)	2.17 (1.44)	4.97 (.97)
All prime conditions	4.13 (2.16)	6.96 (1.95)	6.44 (2.05)	2.13 (1.51)	

over circular explanations, all p s < .001 (see Table 2). Mechanistic and teleological ratings did not differ from each other, $t(411) = .63$, $p = .531$. We take this pattern to reflect chronic explanatory preferences, which form the basic profile on top of which we might expect to see shifts driven by contextual utility.¹ Ratings were also higher under the categorical task than the causal task (Tukey's HSD $p = .039$) and baseline ($p < .001$) conditions; however, this main effect of task has no bearing on the questions we investigate here.

Most importantly, we found a significant interaction between explanation type and task, $F(9, 1224) = 5.73$, $p < .001$, $\eta_p^2 = .040$. A series of planned contrasts supported our prediction that explanation ratings would be boosted in the context of a congruent task. Three separate contrasts compared ratings of formal, mechanistic, and teleological explanations in the context of the congruent task versus the average of ratings for that explanation type in the other two (incongruent) task conditions. As predicted, each explanation type was rated as significantly better under the congruent task compared to the other generalization conditions: formal

explanation $F(1, 408) = 9.85$, $p = .002$, $\eta_p^2 = .024$; mechanistic explanation, $F(1, 408) = 7.36$, $p = .007$, $\eta_p^2 = .018$; teleological explanation, $F(1, 408) = 7.23$, $p = .006$, $\eta_p^2 = .019$ (see Fig. 1). Circular explanations were not significantly influenced by task (one-way ANOVA, $F(3, 408) = 2.48$, $p = .061$).

As a further test of the relationship between explanatory preferences and task context, we classified participants based on the explanation type for which they gave the highest average ratings. Twenty ties (18 between mechanistic and teleological explanations) were excluded. As shown in Fig. 2, the distribution of explanation preferences varied significantly across tasks, $\chi^2(9, N = 392) = 31.87$, $p < .001$. Standardized residuals indicate that the effect was driven by participants being more likely to favor mechanistic and, marginally, teleological explanations within the corresponding congruent task contexts (standardized residuals 2.5, 1.9), and less likely to favor these explanations within incongruent task contexts (standardized residuals -2.3, -2.4). The latter pattern, which suggests competition between cause- and function-based reasoning, was additionally supported by a negative correlation between ratings of mechanistic and teleological explanations, $r(410) = -.19$, $p < .001$. No other pair of explanation ratings was significantly negatively correlated.

The baseline condition was originally included to evaluate whether the perceived quality of explanations of a given type was improved, relative to baseline, in the context of a

¹ The lower ratings for formal explanations could in part be due to uncertainty about the nature of the connection between category membership and the feature in question (principled vs. statistical, where only the former supports formal explanation; Prasada & Dillingham, 2006, 2009b) or due to the need to take an extra step in *inferring* a principled connection between a feature and the kind (whereas participants did not have to make such inferences for mechanistic explanations). We thank an anonymous reviewer for pointing this out.

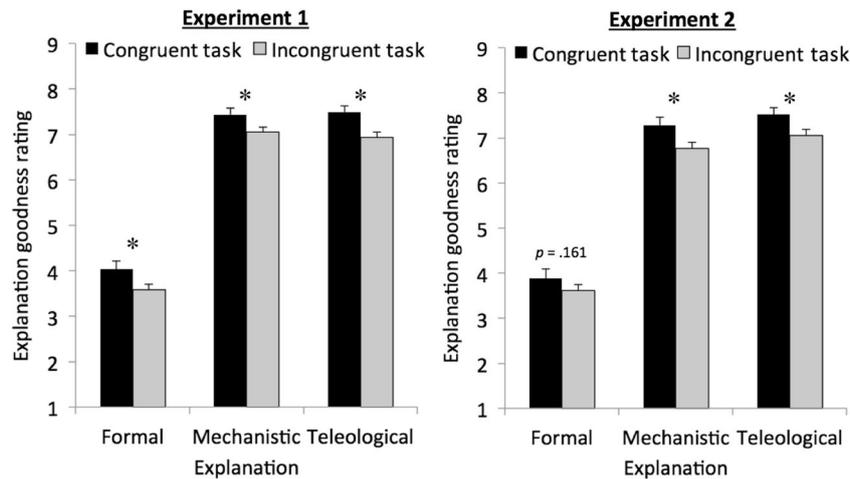


Fig. 1 Explanation goodness ratings as a function of explanation type and task in Experiments 1 and 2; error bars represent 1 SEM; asterisks indicate contrasts significant at $p < .05$

congruent task, or instead depressed, relative to baseline, in the context of an incongruent task. However, the experiments did not, as a whole, support a clear and consistent story; we therefore bracket consideration of this condition (see [Supplementary Materials](#) for details).

In sum, Experiment 1 reveals that contextual utility affects the perceived quality of explanations. Statements that explained an observation in terms of category membership (formal), in terms of proximal causal mechanisms (mechanistic), or in terms of functions (teleological) were perceived as better explanations in the context of tasks that called for the information provided by these explanations. As expected, these effects acted on top of more stable explanatory preferences (favoring mechanistic and teleological explanations over formal explanations), which likely reflect the chronic utility of these explanations across a variety of contexts (see Lombrozo & Rehder, 2012, for a relevant discussion of functional explanations). Finally, these effects were observed even though all participants received exactly the same information about the categorical, mechanistic, and functional relationships involving the explained features.

Experiment 2

Experiment 2 had two objectives. First, to verify the reliability of the small effects observed in Experiment 1, we aimed to replicate the interaction between task and explanation types, but with a different manipulation of contextual utility: Participants were given a task as a museum assistant, which involved classification (grouping items), proximate causes (identifying how something came about), or functions (identifying functions). Second, we aimed to better understand the mechanism underlying the effect of contextual utility on explanations. On van Fraassen’s (1980) account, context can influence an explanation in several ways: by changing the general topic, the contrast class, or the relevance relation. Even when a topic and contrast class are fixed, however, the relevance relation can remain underspecified. For example, if one asks why blood circulates through the body (as opposed to not circulating through the body), either a mechanistic explanation (“because the heart pumps the blood through the arteries”) or a teleological explanation (“to bring oxygen to every part of the body tissue”) would stand in an appropriate

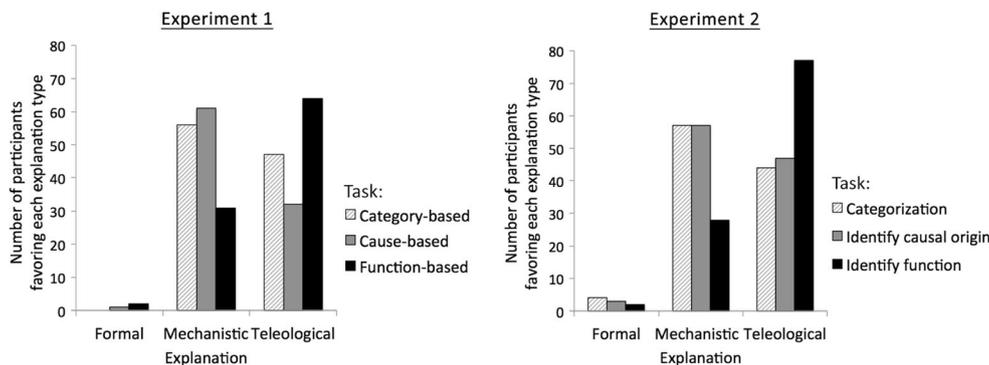


Fig. 2 Explanation preferences as a function of task in Experiments 1 and 2. Circular explanations were only favored by one participant across all experiments, and are not shown

relevance relation to the question, even if the contrast class is fixed to {blood circulates; blood does not circulate} ($p = .142$). Given our interest in effects of contextual utility, we aimed to investigate whether our task manipulation could influence the evaluation of formal, mechanistic, and teleological explanations even when the contrast class of the why-question was explicitly fixed across tasks.

Method

Participants Four-hundred-and-ninety-six participants were recruited on Amazon Mechanical Turk in exchange for \$1.65. An additional 317 participants were excluded for failing a memory check.

Materials, design, and procedure Experiment 2 mirrored Experiment 1, except as noted. First, we introduced a cover story that the participant was a museum assistant who needs to figure out one of three things: how new objects or organisms should be grouped in the museum (categorization task), how objects or organisms come to possess certain properties (causal origin task), or what functions the properties of objects or organisms serve (functional task). The task reinforcers were adapted accordingly (Table 3 illustrates all changes; see also [Supplementary Materials](#) for a sample trial). Second, to test the possibility that effects of task on explanation judgments were produced (only) by a shift in the implied contrast class of the questions, we added a clarification to the explanation probes specifying the contrast class (e.g., “Why does this item lower its temperature when food is about to go bad (as opposed to not lowering it)?”). Finally, domain was manipulated between subjects.²

Results

Explanation ratings were analyzed in an ANOVA, with explanation type as a within-subjects factor and task as a between-subjects factor. The main effect of explanation type was replicated, $F(3, 1476) = 946.06, p < .001, \eta_p^2 = .658$: participants preferred mechanistic and teleological explanations over formal explanations, which were all preferred over circular explanations, $ps < .001$ (see Table 2). Mechanistic and teleological explanation ratings did not differ from each other, $t(495) = 1.64, p = .102$. The task manipulation also produced a significant main effect (irrelevant for our hypothesis), $F(3, 492) = 2.71, p = .004, \eta_p^2 = .016$; it was driven by higher ratings within the categorical task

² Experiment 2 ended with an additional exploratory task that examined whether the effect of task extends to judgments of an explanation’s *probability* in addition to its quality, as might be anticipated if an explanation’s “loveliness” is used as a cue to its “likeliness” (Lipton, 2004). There were no significant effects of explanation manipulation (see [Supplementary Materials](#) for details).

than the functional task (Tukey’s HSD $p = .046$, all remaining $ps \geq .190$).

Most importantly, there was a significant interaction, $F(9, 1476) = 4.00, p < .001, \eta_p^2 = .024$. Planned contrasts showed that mechanistic and teleological explanations were rated significantly higher under the congruent task than the incongruent task (mechanistic explanations: $F(1, 492) = 4.49, p = .035, \eta_p^2 = .009$; teleological explanations: $F(1, 492) = 5.59, p = .018, \eta_p^2 = .011$; see Fig. 1). However, the contrast did not reach significance for formal explanations, $F(1, 492) = 1.97, p = .161$. Ratings of circular explanations were not influenced by task: (one-way ANOVA, $F(3, 492) = .20, p = .898$).

As in Experiment 1, we also found that the distribution of explanation preferences varied as a function of task, $\chi^2(6, N = 433) = 26.19, p < .001$. (This analysis excluded 60 ties, 50 of which were between mechanistic and teleological explanations, evenly spread across conditions.) As shown in Fig. 2, the effect was driven by the functional task, for which fewer participants preferred mechanistic explanations and more preferred teleological explanations (standardized residuals -3.0 and 2.9), as in Experiment 1. In the causal task, differences were in the predicted directions, but did not reach significance (standardized residuals $1.2, -1.2$). Once more, ratings of mechanistic and teleological explanations were negatively correlated, $r(494) = -.19, p < .001$. No other pair of explanation ratings was significantly negatively correlated.

Discussion

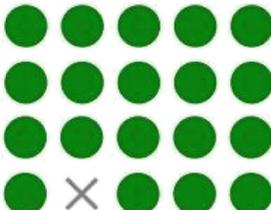
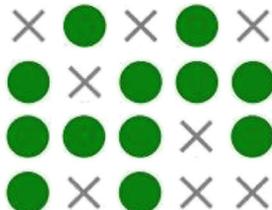
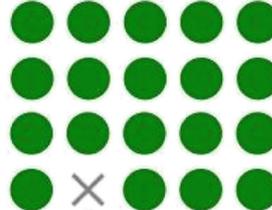
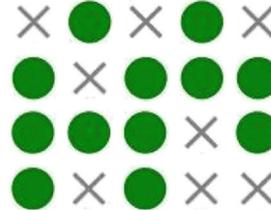
Mirroring Experiment 1, Experiment 2 revealed an interaction between task and explanation ratings: explanations that offered greater contextual utility were rated more highly, with significant effects for mechanistic and teleological explanations, and a matching trend for formal explanations.³ These effects were found with explanation requests that fixed the contrast class across tasks.

³ We can speculate that the weakened boost for formal explanations in Experiment 2 had to do with the indirect nature of the category-based task, for which participants were asked to decide whether two items belonged to the same part of a museum. Even though participants were invited to put items of the same kind together (instructions mentioned that “for example, stores often put objects of the same kind next to each other” or “zoos often group animals of the same kind together”), participants were not explicitly told to organize objects in the museum based on categorical rather than thematic principles. In contrast, in Experiment 1 the manipulation of the contextual utility of categorical relationships was more direct: In the category-based task, participants specifically sought information about the category membership of an item (e.g., “find out if it’s a glenta”) and then used this information to make an inference about that item. The relative vagueness of the contextual utility prime for categorical relationships in Experiment 2 may account for the weakened effect on formal explanation evaluation relative to Experiment 1.

Table 3 Sample instructions, explanations, and task-reinforcers used in Experiment 2. Task and domain were manipulated between subjects; explanation type was manipulated within subjects (with one explanation presented for evaluation on each trial). Text referring to living things is provided in square brackets

1. Instructions (Task manipulation)									
In this experiment, you will be the assistant to a museum director. The museum will present the public with little-known organisms and objects. Your job will be to figure out...									
<p>Task: Categorization ...how these objects [organisms] should be <i>grouped</i> with others in the museum. For example, stores often put objects of the same kind next to each other [zoos often group animals of the same kind together]. Your job will be to figure out how the objects [organisms] should be organized in the museum.</p>	<p>Task: Identify causal origin ...<i>how it is</i> that objects [organisms] come to have certain traits. For example, when engineers encounter novel objects, they may need to “reverse-engineer” them to figure out [botanists and zoologists often try to figure out] what produces some characteristic that they observed in that object [a plant or animal.] Your job at the museum will be to figure out what produces features of objects [living organisms]: how they do certain things, or come to have certain characteristics.</p>	<p>Task: Identify function ...what each object (or some feature of an object) <i>is for</i> [what the biological traits (such as parts or behaviors) of each organism <i>are for</i>.] For example, archaeologists often identify the functions of the objects that they find [biologists often identify the functions of the biological traits of animals or plants that they are studying]. Your job at the museum will be to identify the functions of exhibited objects [organisms].</p>							
2. Description									
	<p>The uniory is a new type of refrigerator. It has a martion sensor, which automatically lowers the temperature when it senses that any food is about to go bad. The lowered temperature keeps the food fresh longer.</p>								
3. Explanation evaluation									
<p>Below is a picture of one particular item, #JX8Y-d9921, from a retail facility. Why does this item <i>lower its temperature when food is about to go bad?</i> (as opposed to not lowering its temperature when food is about to go bad)</p>									
<p>Formal explanation: Because it’s a uniory, and uniories lower their temperature when food is about to go bad.</p>	<p>Mechanistic explanation: Because it has a martion sensor, which lowers the temperature when food is about to go bad.</p>	<p>Teleological explanation: Because lowering the temperature when food is about to go bad keeps the food fresh longer.</p>	<p>Circular explanation: Because some things lower temperature when food is about to go bad.</p>						
Very bad explanation (1) – Very good explanation (9)									
4. Task reinforcement									
<p>Now you receive two completely new objects: item A and item B. Each one may or may not be a uniory. Both of them lower their temperature when food is about to go bad.</p>									
<p>Task: Categorization Do you think items A and B both belong in the <i>same part of the museum?</i></p>	<p>Task: Identify causal origin Do you think the <i>same factor</i> produces this characteristic in both item A and item B? Definitely no (1) - Definitely yes (9)</p>	<p>Task: Identify function Do you think this characteristic serves the <i>same function</i> in item A and item B?</p>							
		<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 80%;">Item</th> <th style="width: 10%;">A</th> <th style="width: 10%;">B</th> </tr> </thead> <tbody> <tr> <td>Lowers its temperature when food is about to go bad?</td> <td style="text-align: center;">Yes</td> <td style="text-align: center;">Yes</td> </tr> </tbody> </table>		Item	A	B	Lowers its temperature when food is about to go bad?	Yes	Yes
Item	A	B							
Lowers its temperature when food is about to go bad?	Yes	Yes							

Table 4 Sample priming procedure used in Experiment 3: causal prime. The type of prime (inductive utility or salience) and target feature (cause or function) were manipulated between subjects. The left–right position of options was randomized for each item, for each participant. The feedback patterns were selected from a pool of pairs of one high (90%–95%) and one low (55%–60%) proportion of green dots

Inductive utility prime: “Guessing game”	Salience prime: “Shopping game”
<div data-bbox="146 283 263 415" style="float: left; margin-right: 10px;"> </div> <p>1. Description The uniory is a new type of refrigerator. It has a martion sensor, which automatically lowers the temperature when it senses that any food is about to go bad. The lowered temperature keeps the food fresh longer.</p>	
<p>2. Choice + Answer Select a hint: Behind this box there is a refrigerator. You need to guess if it lowers its temperature when food is about to go bad. Choose to reveal one piece of information: does it have a martion sensor? [reveal] does it keep food fresh longer? [reveal] Provided answer: Yes / No (determined randomly)</p>	<p>2. Guess What do you think the average cost of a new uniory is? \$1100 or less More than \$1100 - Thank you for your guess. You won’t be told whether you were right or wrong.</p>
<p>3. Feedback Rate of successful guesses after picking this hint:  Rate of successful guesses after picking this hint:  - Thank you for your guess. You won’t be told whether you were right or wrong.</p>	<p>3. Choice Imagine you are buying a refrigerator. An online seller describes the product with several features, including the two features below. Select the feature that you think is <i>harder to remember</i>. Has a martion sensor [select] Keeps food fresh longer [select]</p>
<p>4. Guess:  Does the refrigerator behind the box lower its temperature when food is about to go bad? No / Yes</p>	<p>4. Feedback Thank you for your choice. Proportion of shoppers who chose this option when it was offered: Proportion of shoppers who chose this option when it was offered:  </p>

Experiment 3

Experiments 1 and 2 support our primary prediction: The perceived quality of an explanation is affected by contextual utility. However, there are multiple ways to interpret this result.

One possibility is that it reflects a beneficial feature of our capacity to assess explanations. By favoring explanations that provide task-relevant information, participants could be privileging relationships with high inductive utility, where the calculation of inductive utility is calibrated to context. In

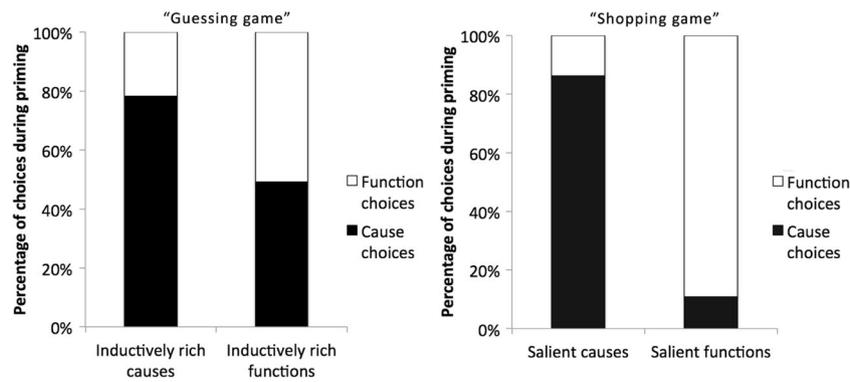


Fig. 3 Percentages of selected cause and function features by primed feature type, in inductive utility (*left*) and salience (*right*) priming conditions

Experiment 3, we test this hypothesis by manipulating inductive utility via direct feedback: We provide participants with explicitly stated information about which kinds of features (causes or functions) are inductively useful, with the prediction that the corresponding explanations will receive a boost in perceived quality.

Experiment 3 also tests two alternative possibilities. First, it could be that the influence of task on explanation ratings reflects an effect of *mere salience*. For instance, drawing attention to some features could increase the fluency of processing explanations that contain those features, resulting in higher ratings. To investigate whether mere feature salience is sufficient to drive effects of task on explanation ratings, Experiment 3 also includes conditions in which causal or functional features are made salient, but not because they are inductively useful.

A second possibility is that effects of task on explanation ratings are driven by perceived *task demands*: Our task manipulations could be taken as a cue to the “correct” explanation the experimenter intends. To evaluate this possibility, we

include a posttest asking participants what the experiment is about, and we analyze performance as a function of their assumptions.

Experiment 3 thus includes a total of four priming conditions in addition to a no-prime control, the result of crossing prime type (inductive utility vs. salience) with primed feature (cause vs. function). We focus on the contrast between mechanistic and functional explanations, dropping a manipulation of categorical relations and analyses of formal explanations, both to simplify the design and to maximize our chances of finding effects, as both Experiments 1 and 2 suggested more reliable effects of task on these explanation types.

Method

Participants Two-hundred-and-forty-six participants were recruited on Amazon Mechanical Turk in exchange for \$2.25. An additional 35 participants were excluded for failing a memory check.

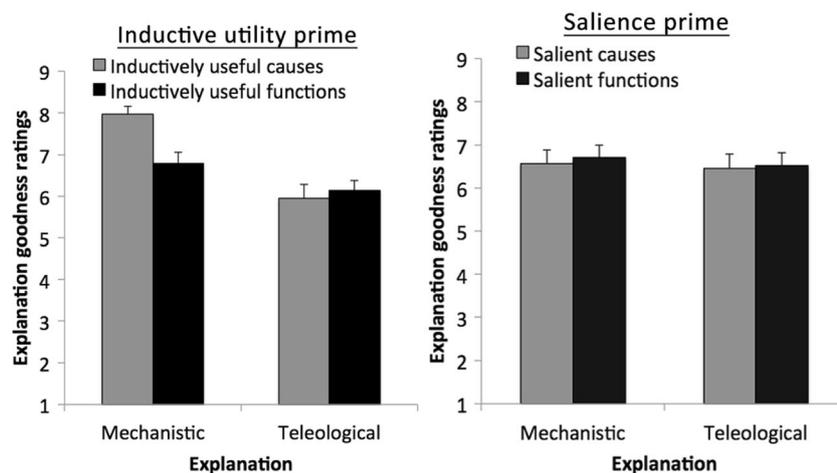


Fig. 4 Explanation goodness ratings as a function of explanation type and target feature in the inductive utility and salience priming conditions. Error bars represent 1 SEM. The no prime condition is not shown (see Table 2 for the means and SDs; and see [Supplementary Materials](#) for more detail)

Materials, design, and procedure Participants were randomly assigned to one of five conditions: a no-prime control, or one of four priming conditions that formed a 2×2 design: prime (inductive utility vs. salience) \times target feature (cause vs. function).

To reduce the number of conditions, Experiment 3 focused on artifacts, using the same eight descriptions from Experiments 1 and 2. Participants first read the descriptions during a priming phase, and then completed an explanation evaluation phase. In each phase, the order of items was randomized for each participant.

In the control condition, participants were given instructions to simply read the eight descriptions, without performing another task. In the inductive utility conditions, participants played a “guessing game” in the course of which they learned that either causes or functions had high predictive validity. For example, after learning about a refrigerator that has a *martion sensor that lowers the temperature in order to keep the food fresh longer*, participants were told that they would need to guess whether a new refrigerator lowers its temperature (i.e., possesses the *middle feature* in the causal chain). They could then choose to learn one fact about the new refrigerator before making their guess, specifically whether it *has the martion sensor (cause)* or *keeps the food fresh longer (function)*. After choosing, they received yes-or-no feedback regarding whether the object possessed the feature in question (in equal proportion, randomly assigned), and, crucially, they also received feedback about the rate of successful guesses after picking each cue, presented in the form of dot diagrams (see Table 4 for an example and [Supplementary Materials](#) for a sample trial). Across eight priming trials, either cause or function features were consistently indicated as having higher predictive utility. After the predictive utility feedback, participants made their guess about the new refrigerator, but were not told whether they had guessed correctly. At the end of the priming phase, participants were told that they would play another guessing game at the end of the experiment, in order to sustain the relevance of the inductive cues from the priming phase.

In the salience priming conditions, participants played a “shopping game” in the course of which either cause or function features were strongly emphasized (see Table 4 and [Supplementary Materials](#)). Matching the inductive utility primes, on each trial participants read a description, made one guess, and chose either a cause or a function feature. Cause or function features were made salient by capitalizing on the fact that most cause features involved novel terms (martion sensor, nordrum part, etc.); we could therefore draw attention to cause features by asking participants to select features that were *harder* to remember, or draw their attention to function features by asking them to select features that were *easier* to remember. (A pretest with 77 participants showed that asking people to select features that are harder vs. easier to remember reliably

encouraged people to select either cause or function features, $t(37) = 14.82$, $p < .001$, Cohen’s $d = 4.79$, and also found that participants tended to have better memory for which features were primed in the salience condition than in the inductive utility condition, $t(75) = 2.70$, $p = .009$, Cohen’s $d = .61$, suggesting the prime was quite effective). To further reinforce the salience of causes or functions, participants received feedback indicating that the primed feature type was consistently selected by the majority of participants. Critically, however, the feature choice (regarding memorability) was not inductively relevant to the guess (regarding product cost); thus salience was manipulated independently of inductive utility. At the end of the priming phase, participants were told that they would play another shopping game at the end of the experiment.

For the explanation evaluation phase, all participants evaluated mechanistic and teleological explanations as in Experiments 1 and 2 (but without interspersed task reinforcers). Next, as promised, participants in priming conditions played a short guessing or shopping game (not analyzed).

Finally, to address the possibility that participants were responding to task demands, at the end of the experiment we asked them to guess what “this study was getting at.” We provided seven potential research questions in random order (listed in Fig. 6) and asked participants to rate how plausible they found each one, from zero (*not plausible at all*) to 100 (*very plausible*).⁴

Results

Attention manipulation check We first verified that the salience primes were at least as successful as the inductive utility primes in drawing attention to the target feature. We compared the percentage of times participants chose causal features as a function of prime type (inductive utility vs. salience) and target feature (cause vs. function) in a 2×2 ANOVA. The analysis revealed significant main effects of prime type, $F(1, 195) = 19.55$, $p < .001$, $\eta_p^2 = .091$, and target feature, $F(1, 195) = 230.95$, $p < .001$, $\eta_p^2 = .542$, as well as a significant interaction, $F(1, 195) = 44.81$, $p < .001$, $\eta_p^2 = .187$. As shown in Fig. 3, participants chose cause features significantly more often when they were primed than when they were not, in both the inductive utility, $t(99) = 4.96$, $p < .001$, Cohen’s $d = .99$, and salience, $t(96) = 21.93$, $p < .001$, Cohen’s $d = 4.43$, conditions. Moreover, the effect was *larger* for the salience primes. These findings suggest the primes succeeded in drawing

⁴ At the end of the experiment (but prior to the research question plausibility ratings) participants completed an additional task, in which we asked them to generalize the middle feature to new items based on a shared causal feature, function feature, or category label. Our intention was to investigate whether effects of the prime and target features would extend to novel generalizations. However, we did not find such an effect (three-way interaction $F(2, 390) = .55$, $p = .576$), likely in part due to ceiling effects; hence, we do not discuss these results further.

attention to the intended features, and if anything, that the salience condition was more effective in doing so. Having verified that the salience manipulation succeeded in drawing attention to the target features, we next present results from the main task.

Inductive utility versus salience We first tested whether the effects of inductive utility differed from those of mere salience. We performed a mixed ANOVA with explanation type (mechanistic, teleological) as a within-subjects factor and target feature (cause, function) and prime type (inductive utility, salience) as between-subjects factors. The predicted three-way interaction was marginal, $F(1, 195) = 2.77, p = .098, \eta_p^2 = .014$, with effects in the predicted direction. We therefore ran separate analyses for the inductive utility and salience conditions, examining the relationship between primed relation (cause-based, function-based) and the evaluation of explanations.

Effects of inductive utility To test whether the inductive utility of an explanatory relationship boosts the perceived quality of corresponding explanations, we ran a mixed ANOVA with target feature as a between-subjects factor and explanation type as a within-subjects factor. Both target feature, $F(1, 99) = 4.38, p = .039, \eta_p^2 = .042$, and explanation type, $F(1, 99) = 20.80, p < .001, \eta_p^2 = .174$, affected explanation evaluations. As shown in Fig. 4, both of these effects were qualified by a significant interaction, $F(1, 99) = 5.47, p = .021, \eta_p^2 = .052$: mechanistic explanations were rated significantly higher when they were inductively useful than not (planned contrast $p = .001$); but ratings of teleological explanations were not affected by the inductive utility prime (planned contrast $p = .656$). These findings suggest that explanations containing inductively useful relationships are rated more highly, but the effect may depend on the kind of relationship.

We also analyzed the distribution of explanation preferences: The effect of target feature was not significant, $\chi^2(1, N = 82) = .87, p = .352$, but trends were in the predicted directions (see Fig. 5). As in Experiments 1 and 2, we found a significant negative correlation between ratings of mechanistic and teleological explanations, $r(101) = -.21, p = .039$.

Effects of mere salience To test whether mere salience boosts the perceived quality of corresponding explanations, we ran a 2 (target feature) \times 2 (explanation type) mixed ANOVA, which showed no significant effects (explanation type): $F(1, 96) = .21, p = .645$; target feature: $F(1, 96) = .11, p = .741$; interaction: $F(1, 96) = .01, p = .911$). These findings suggest that merely making an explanatory feature salient is not sufficient to boost the perceived quality of corresponding explanations. The effect of target feature on the distribution of explanation preferences (see Fig. 5) was also not significant, $\chi^2(1, N = 82) = .06, p =$

.803, nor was the correlation between ratings of mechanistic and teleological explanations, $r(98) = -.09, p = .383$.

Plausibility ratings Participants' average plausibility ratings for the seven potential research questions are shown in Fig. 6. Ratings did not vary as a function of condition: Question \times Prime, $F(6, 1128) = .59, p = .743$; Question \times Target Feature, $F(6, 1128) = 1.66, p = .127$, three-way interaction, $F(6, 1128) = 1.20, p = .302$. The option most accurately describing the inductive utility condition, labeled (a) in Fig. 6, was not among the most highly-rated questions, and its ratings did not correlate with the mean difference between mechanistic and teleological ratings in any condition (all $ps \geq .118$), suggesting that our findings were not a product of perceived task demands. The research question most accurately describing the salience condition, labeled (b) in Fig. 6, also did not correlate with ratings in any condition (all $ps \geq .172$).

Discussion

Experiment 3 asked three questions. First, does manipulating the inductive utility of a feature type (cause vs. function) through feedback influence the perceived quality of explanations containing that feature (mechanistic vs. teleological)? The answer is “yes”: The inductive utility primes succeeded in boosting the quality of the primed explanation type relative to its alternative, although the effect was only significant for mechanistic explanations.⁵ Second, is making a particular feature type *salient* sufficient to influence the perceived quality of explanations containing that feature? The answer is “no”: While our manipulation of salience successfully influenced choices (indeed, it did so more successfully than the inductive utility manipulation), it did not influence explanation ratings. Third, are shifts in explanation ratings a consequence of participants' inferences about the experimenters' expectations—that is, are they an artifact of task demands? Again, the answer is “no”: Participants were not especially skilled in guessing the true aims of the study, and their guesses did not predict performance.

Taken together with Experiments 1 and 2, these findings suggest that the perceived quality of an explanation depends on its contextual utility: Explanations seem better when they are

⁵ We can only speculate why in Experiment 3 this effect was not significant for ratings of teleological explanations (in fact, priming the inductive utility of functions marginally *suppressed* ratings of teleological explanations relative to the no-prime baseline). In the “guessing game” (the inductive utility priming block) the overall preference was towards selecting causal features, and priming the inductive utility of functions resulted in a roughly equal proportion of function and cause feature choices (see Fig. 3). This may reflect that participants were conflicted about the relative inductive value of cause versus function features. This conflict might have exacerbated the competition between mechanistic and teleological explanations (observed in all three experiments), producing an overall penalty for teleological explanations. Future work is needed to evaluate this speculative explanation.

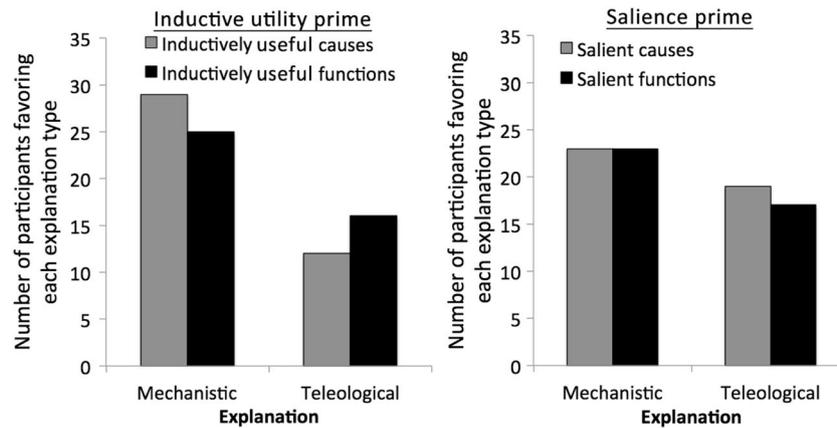


Fig. 5 Explanation preferences as a function of primed feature and explanation type in Experiment 3. Formal explanations, not shown, were preferred over all other explanation types by 4% of participants.

The analysis excluded 11 ties from the inductive utility condition and 14 ties from the salience condition

underwritten by generalizations that are inductively useful in the context of a given task. Experiment 3 also offers preliminary evidence against an alternative explanation of our results in terms of mere salience. However, the results of Experiment 3 should be interpreted with some caution: it's likely that explanation ratings are subject to other performance errors, and the statistical difference between effects of inductive utility and those of salience was marginal. Moreover, salience is likely to be an inductively relevant cue in many real-world contexts; in the present experiment, we took care to dissociate salience from inductive utility.

General discussion

Across three studies involving different manipulations of the contextual utility of explanations, we found that people prefer explanations that highlight the kinds of relationships that they expect to be useful for the task at hand. This was the case for formal, mechanistic, and teleological explanations in

Experiment 1, for mechanistic and teleological explanations in Experiment 2, and for mechanistic explanations in Experiment 3. The reported effects were small but reliable, and for the most part driven by a large proportion of participants (as suggested by the analyses of explanation preferences). Not surprisingly, all three studies also supported the existence of relatively stable explanatory preferences, with mechanistic and teleological explanations rated reliably better than formal explanations, which were in turn better than circular explanations. These baseline preferences could reflect the global inductive utility of each explanation type across habitual contexts, with local contextual utility having a small but systematic effect on top of these general preferences.

Importantly, we found that embedding explanations within different tasks did not simply shift the implied contrast class for an explanation request (which was specified in Experiment 2), but instead affected the relative ratings for different *kinds* of explanations, with task-congruent explanations receiving a relative boost. We also found in Experiment 3 that neither

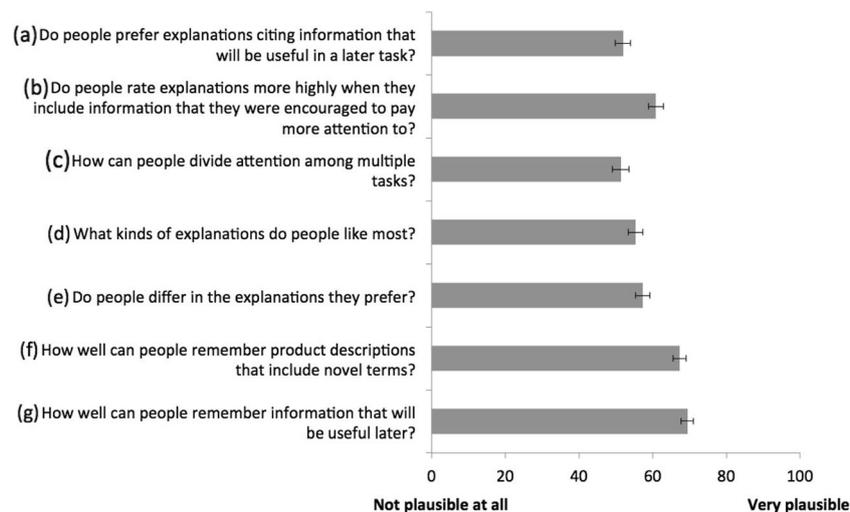


Fig. 6 Plausibility ratings of potential research questions; error bars represent 1 SEM

variations in the mere salience of features nor intuitions about experimenters' expectations can account for our findings. Instead, it appears that effects of contextual utility on explanation evaluation are driven by the inductive value of different kinds of explanatory relationships, consistent with the Explanation for Export proposal stating that good explanations supply information with high anticipated utility (Lombrozo & Carey, 2006).

These findings also have implications for philosophical accounts of explanation. One of the main critiques of pragmatic accounts is the lack of constraint on the relation between candidate explanations and what they explain (Kitcher & Salmon, 1987). Our work demonstrates that the task pursued by the explainer can systematically constrain that relation, which raises the possibility of a pragmatic approach that is appropriately constrained and descriptively adequate as an account of human judgments. That said, our findings do not rule out more traditional accounts of explanation. For instance, accounts that allow for incomplete (Hempel & Oppenheim, 1948) or partial explanations (Kitcher, 1989; Railton, 1978) could accommodate our results if our manipulation impacted which parts of the “complete” explanation were selected (but see Woodward, 2003). Alternatively, our results could be accommodated by allowing for pluralism in the patterns, covering laws, or other structures governing explanations, with contextual utility fixing the structure with respect to which explanations are evaluated at a given time.

Our findings also provide potential evidence for competition between mechanistic and function-based reasoning (see also Heussen, 2010; Lombrozo & Gwynn, 2014). In Experiment 1, teleological explanations were rated significantly lower under the cause-based task compared to other generalization conditions, and in Experiment 2, mechanistic explanations were rated significantly lower under the functional task relative to other generalization conditions, suggesting that in addition to boosting task-congruent explanations, contexts can also penalize task-incongruent explanations. Notably, this pattern of competition was restricted to mechanistic versus function-based reasoning: only the causal and functional tasks produced suppression effects, and only ratings of mechanistic and teleological explanations were significantly negatively correlated.

Relationship to prior work

Our findings are consistent with prior work suggesting a close relationship between explanation and inference. For example, Lombrozo and Gwynne (2014) and Vasilyeva and Coley (2013) found that different types of explanations predicted different patterns of property generalization (for similar effects in categorization, see Ahn, 1998; Lombrozo, 2009). These studies, however, did not investigate a relationship in the reverse direction, with (anticipated) inferences affecting explanation judgments.

Prior work also suggests that the production of teleological and mechanistic explanations can depend on context (Chin-Parker & Bradner, 2010; Hale & Barsalou, 1995), although these studies manipulated context quite differently from the studies reported here: specifically, they varied background conditions and participants' knowledge about what they observed (for instance, whether the outcome to be explained was accidental and idiosyncratic versus intended and systematic). To our knowledge, our studies provide the first demonstration that contextual utility can affect the perceived quality of explanations even when participants' background knowledge is held constant. In our studies, the relationships underlying the formal, mechanistic, and functional explanations always held; what varied was the contextual utility of that relationship, and the perceived quality of the corresponding explanation.

Our findings differ from those of Chin-Parker and Bradner (2010) in that they found effects of context on explanation *generation*, but not on explanation *evaluation*. We speculate that such effects were not found in their studies, but emerged in ours, due to methodological differences.⁶ Overall, though, we agree with Chin-Parker and Bradner (2010) that “the constraints inherent within the tasks of evaluating and generating explanations are not equivalent” (p. 230). We would expect a partial overlap between such constraints, and anticipate valuable insights coming from a systematic investigation of shared versus unique constraints, using a range of tasks and experimental paradigms.

Future directions and conclusion

Our findings demonstrate that contextual utility can affect the perceived quality of different kinds of explanations, and that this is unlikely to be a product of low-level attentional mechanisms or intuitions about experimenters' expectations. However, further work is needed to specify the scope and basis of this effect. With respect to scope, are people responsive to the contextual utility of an explanation for its intended *recipient*, or are they restricted to evaluating contextual utility from their own perspective? What are relevant markers of inductive utility, beyond past and anticipated inferences? (Salience may well turn out to be one such marker; we took

⁶ Some methodological differences between our studies and those of Chin-Parker and Bradner (2010) may account for the fact that they did not observe the effects of context on explanation *evaluation* reported here: First, our participants answered well-defined why-questions, whereas participants in Chin-Parker and Bradner's studies received a broad request to explain “what [they] just saw” in a video clip. Second, their participants always evaluated explanations after generating explanations and completing an explanation selection task where they were “given the contrast class and relevance relation, instead of being asked to construct [...] these things” (p.244). Subsequent explanation ratings may well have been influenced by these preceding tasks. Thus, it is not surprising that we find contextual influences where Chin-Parker and Bradner (2010) did not. More generally, we would expect effects of context and contextual utility to be quite widespread—Hilton and Erb (1996), for example, do report significant effects of shifting the contrast class on explanation evaluation.

special care to disentangle salience from inductive utility in Experiment 3, but in the real world there may exist a correlation between them that reasoners can exploit.) Finally, what are the mechanisms that underlie these effects? Context could plausibly influence memory, categorization, information search, the conscious or unconscious selection of alternatives, and much more (see Aarts & Elliot, 2012; Dijksterhuis & Aarts, 2009, for reviews)—which in combination could be said to induce different stances (Dennett, 1987).

Identifying the psychological processes that contribute to effects of contextual utility is an important question for future research. Although much work remains to be done, our studies take an important step towards developing a psychological account of explanation that recognizes the context-sensitive and flexible nature of human explanatory judgments.

Acknowledgements This work was supported by the Varieties of Understanding Project, funded by the John Templeton Foundation. We are grateful to Daria Serrano Cargol, Siqi Liu and Marc Collado-Ramírez for assistance in preparation of materials and data collection.

References

- Aarts, H., & Elliot, A. (Eds.). (2012). *Goal-directed behavior*. New York, NY: Taylor & Francis.
- Ahn, W. (1998). Why are different features central for natural kinds and artifacts? *Cognition*, 69, 135–178.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11, 211–227.
- Chin-Parker, S., & Bradner, A. (2010). Background shifts affect explanatory style: How a pragmatic theory of explanation accounts for background effects in the generation of explanations. *Cognitive Processing*, 11(3), 227–249.
- Craik, K. J. W. (1943). *The nature of explanation*. Cambridge, UK: Cambridge University Press.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un) consciousness. *Annual review of psychology*, 61, 467–490.
- Gorovitz, S. (1965). Causal judgments and causal explanations. *The Journal of Philosophy*, 62(23), 695–711.
- Hale, C. R., & Barsalou, L. W. (1995). Explanation content and construction during system learning and troubleshooting. *Journal of the Learning Sciences*, 4(4), 385–436.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: John Wiley & Sons.
- Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175.
- Heussen, D. (2010). When functions and causes compete. *Thinking & Reasoning*, 16(3), 233–250.
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, 107(1), 65–81.
- Hilton, D. J., & Erb, H.-P. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking & Reasoning*, 2(4), 273–308.
- Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies. Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, 142(4), 1074–1083.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation*. Minneapolis: University of Minnesota Press.
- Kitcher, P., & Salmon, W. (1987). Van Fraassen on explanation. *Journal of Philosophy*, 84, 315–330.
- Leake, D. (1995). Abduction, experience, and goals: A model of everyday abductive explanation. *Journal of Experimental & Theoretical Artificial Intelligence*, 7(4), 407–428.
- Lipton, P. (1990). Contrastive explanation. *Royal Institute of Philosophy Supplement*, 27, 247–266.
- Lipton, P. (1993). *Is the Best Good Enough?* Proceedings of the Aristotelian Society, 93, new series, 89–104.
- Lipton, P. (2004). *Inference to the best explanation*. London, UK: Psychology Press.
- Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, 55, 232–257. doi:10.1016/j.cogpsych.2006.09.006
- Lombrozo, T. (2009). Explanation and categorization: How “why?” informs “what?”. *Cognition*, 110(2), 248–253.
- Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99(2), 167–204.
- Lombrozo, T., & Gwynne, N. Z. (2014). Explanation and inference: mechanistic and functional explanations guide property generalization. *Frontiers in Human Neuroscience*, 8, 700.
- Lombrozo, T. & Rehder, B. (2012). Functions in biological kind classification. *Cognitive Psychology*, 65, 457–485. doi:10.1016/j.cogpsych.2012.06.002
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129(4), 592.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology*, 57(2), 189.
- Patterson, R., Operskalski, J. T., & Barbey, A. K. (2015). Motivated explanation. *Frontiers in human neuroscience*, 9. doi:10.3389/fnhum.2015.00559
- Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, 99(1), 73–112. doi:10.1016/j.cognition.2005.01.003
- Prasada, S., & Dillingham, E. M. (2009a). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33(3), 401–448.
- Prasada, S., & Dillingham, E. M. (2009b). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33(3), 401–48. doi:10.1111/j.1551-6709.2009.01018.x
- Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, 49(1), 123–163.
- Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science*, 45(2), 206–226.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- van Fraassen, B. (1980). *The scientific image*. Oxford, UK: Clarendon Press.
- Vasilyeva, N., & Coley, J. C. (2013). Evaluating two mechanisms of flexible induction: Selective memory retrieval and evidence explanation. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Cognitive Science Society: Austin, TX.
- Vlach, H. A., & Noll, N. (2016). Talking to children about science is harder than we think: Characteristics and metacognitive judgments of explanations provided to children and adults. *Metacognition and Learning*, 11(3), 317–338. doi:10.1007/s11409-016-9153-y
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.