COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70101
© 2025 Cognitive Science Society LLC.
ISSN: 1551-6709 online
DOI: 10.1111/cogs.70101

Seeking the Category: The Pragmatic Function of Formal Explanations and the Role of Cognitive Reflection

Ivan Aslanov,[a] Alexey Kotov,[b] Ernesto Guerra,[a] Alina Fedoriaieva,[b] Tatyana Kotova[c]

[a]*Center for Advanced Research in Education, Institute of Education, Universidad de Chile*
[b]*Laboratory for Cognitive Research, HSE University*
[c]*Department of Psychology, The Russian Presidential Academy of National Economy and Public Administration*

Received 25 November 2024; received in revised form 14 July 2025; accepted 16 July 2025

Abstract

Formal explanations are statements that explain properties of an object by referring to its category. This study investigates the role of pragmatics in the evaluation of formal explanations. Across six experiments, we examined how a questioner's knowledge of category identity and an explanation's capability to specify a category affect satisfaction with such explanations. Experiments 1a and 1b demonstrate that participants find formal explanations less satisfactory when the questioner is already aware of the category identity. Experiments 2a and 2b show that participants assumed a questioner was unaware of an object's category if they were satisfied with the formal explanation. In Experiment 3, open-ended responses revealed that satisfied questioners were perceived as seeking to learn a category identity, while dissatisfied ones were assumed to have other motives. Finally, Experiment 4 compares tautological formal explanations (where a label points to all categories possessing a particular feature at once) and nontautological ones (where a label points to one of several competing categories), and examines the role of cognitive reflection in their evaluation. It demonstrates that people with high cognitive reflection are more sensitive to pragmatic context and value a formal explanation more if it can identify a specific category. This study shows that formal explanations are satisfactory when they fulfill a specific pragmatic function, namely, helping to define a category when the questioner knows only its feature. It also shows that people prone to automatic intuitive responses are less likely to consider this function and tend to evaluate formal explanations independently of this part of the pragmatic context.

Keywords: Formal explanations; Pragmatics; Cognitive reflection; Explanations

Correspondence should be sent to Ivan Aslanov, Center for Advanced Research in Education, Institute of Education, Universidad de Chile, Periodista José Carrasco Tapia 75, Santiago 8330014, Chile.
E-mail: ivaslanov@gmail.com

1. Introduction

Formal explanations are statements that explain the presence of a feature in an object by appealing to the category to which the object belongs. For example, "*this animal flies because it is a bird*" is a formal explanation, as it contains no information beyond the categorical label. The term "formal" traces back to Aristotle's philosophy, where he identified four causes that could underlie any explanation. So, if we want to explain why a match burns, we might say it is made of wood (material cause), or that it illuminates a dark room (function), or that someone lit it (source of action). We could also respond, "because it is a match," which would be a formal explanation, meaning an explanation that appeals to the special nature of all "matches" distinguishing them from all nonmatches.

As demonstrated in the series of works by Prasada and collaborators (Prasada & Dillingham, 2006; Prasada & Dillingham, 2009; Haward, 2021; Prasada, 2021; Rivera, Prasad, & Prasada, 2023), formal explanations aid in the construction of kind representations: they assert that an object possesses certain properties because it belongs to a category for which those properties are expected. However, these explanations are selective and do not apply to all features of an object, even if those features are common among all members of the category. For example, the explanation "*because it is a dog*" would work better for the question "*why does this animal have four legs*" than for "*why does this animal have a collar*," even though dogs often wear collars, and four-legged animals are common across other biological species as well (Prasada & Dillingham, 2009). In this sense, the adequacy of formal explanations may support essentialism as an inherent feature of human cognition (Gelman, 2004; Newman & Knobe, 2019).

Nonetheless, the status of formal explanations is not so clear. Are they truly "real" explanations? It is well-known that the main functions of explanations involve the accumulation of new information, integrating it with prior knowledge, and creating normative expectations about previously unknown phenomena (Lombrozo, 2006; Lombrozo & Carey, 2006; Foster-Hanson & Lombrozo, 2022). It is not entirely clear whether formal explanations can fulfill these functions simply by pointing to an object's category. Some researchers suggest that formal explanations are, in fact, implicitly causal: "If formal explanations were devoid of causal information […] They would appear to us as circular restatements of the phenomena in question" (Spaulding, 2023, p. 639).

Alternative views on the persuasiveness of formal explanations are related to their utilitarian functions, which are not directly connected with explanation per se but serve as certain "secondary benefits." First, formal explanations may indicate the existence of potential additional knowledge within a community: if there is a commonly accepted word to describe a group of phenomena, additional information about it may be discovered (Hemmatian & Sloman, 2018). Second, a label in a formal explanation can act as a placeholder for a proper explanation and suggest that the real cause is hidden within the nature of the object and may be uncovered through research (Gelman, Cimpian, & Roberts, 2018). Third, in some cases, the presence of a specific verbal label can serve as an additional marker that is considered when categorizing a new phenomenon: for instance, if strange behavior is given a scientific-sounding name, people are more likely to consider that this behavior is due to a mental disorder (Aslanov, Sudorgina, & Kotov, 2022; Giffin, Wilkenfeld, & Lombrozo, 2017).

It should be noted that there is some evidence against the claim that the persuasiveness of formal explanations is solely based on their "secondary benefits." For instance, Rivera et al. (2023) provide several experimental findings suggesting that formal explanations are not placeholders but can be explanatory simply by appealing to category membership. They demonstrated this by comparing formal explanations (e.g., "*That flies because it is a bird*"), inherent explanation placeholders (e.g., "*That flies because of deeper internal features*"), and tautological explanations (e.g., "*That flies because it flies*"). Their data suggest that formal explanations are perceived as more naturally sounding, but not necessarily more satisfactory, compared to inherent explanation placeholders. At the same time, this does not necessarily refute the hypothesis made by Gelman et al. (2018); it is possible that the linguistic form of the formal explanation is better suited to indicate certain "deeper internal features" than the phrase "because of deeper internal features" (the understanding of which may require, in some sense, philosophical thinking). Moreover, formal explanation may point to "internal features" of a specific type (for instance, those that condition the mechanism behind the feature), while the explanations from the "inherent explanation placeholders" condition do not clarify which type of "internal features" is meant.

Thus, to understand how and why people use formal explanations, more attention should be given to how and under what circumstances they interpret them. In other words, studying the pragmatic factor of formal explanation use may provide important insights into understanding their nature.

2. Why pragmatics is important for the evaluation of explanations

As noted by Van Fraassen (1988), explanation always differs from a proposition or an argument because it is a response and is inseparably tied to the context. The pragmatic-rhetorical theory of explanation suggests that the persuasiveness and appropriateness of an explanation are determined by the pragmatics of the situation (Faye, 2007). Philosophical and linguistic analysis of explanations indeed shows that people should aim to provide the questioner with information they likely do not possess (Tschaepe, 2009). A growing body of empirical evidence also points to how implicit pragmatic factors, usually not considered by researchers in their experiments, influence participants' evaluation of explanations. For example, depending on background knowledge, people may interpret the same explanation request differently, which affects the type of explanation they choose to generate (Chin-Parker & Bradner, 2010). Vasilyeva, Wilkenfeld, and Lombrozo (2017) demonstrated that the assessment of an explanation's satisfactoriness depends on how well it aligns with the task people are set to solve. They showed that, for example, a functional explanation of a feature in a fictional animal will seem more satisfying if participants are expected to infer the presence of a similar feature in another fictional animal based on functional information. In other words, an explanation like "Why does the *glenta* rise to the ocean's surface during the day? Because that helps it replenish oxygen reserves" will appear more satisfying if participants know that they will later be asked a question such as "This animal also needs to replenish oxygen reserves. Do you think it rises to the ocean's surface during the day?".

By taking pragmatics into account when analyzing the process of generating and evaluating explanations, it becomes possible to offer alternative interpretations of well-known effects. For example, this relates to people's tendency to prefer teleological explanations in many domains, even in those where teleological explanations are clearly inappropriate, such as natural phenomena (Banerjee & Bloom, 2015; Kelemen, 1999; Schachner, Zhu, Li, & Kelemen, 2017). Indeed, natural phenomena occur randomly and not for any specific purpose, and the fact that people tend to view teleological explanations as suitable for explaining nature seems like an irrational bias. This has led researchers to hypothesize about humans' intuitive theism (Kelemen, 2004) or our natural preference to attribute agency to any events or phenomena in the world (Rose, 2020; Rose & Schaffer, 2017). However, as shown by Joo, Yousif, and Keil (2022), considering pragmatics adds a new dimension to understanding these effects. According to their findings, participants have implicit beliefs about the kind of information the questioners expect to receive when evaluating the satisfactoriness of explanations. Therefore, they often choose teleological explanations not because they believe they contain causal information, but because they expect that the questioner is interested in the possible "purpose" of the object. This means that when answering the question "why," they imply that the questioner meant "for what purpose." Thus, Joo et al. (2022) experiments suggest that people are not always as teleological as previously thought. Similarly, if we pay attention to how people interpret situations in which formal explanations are given, we can look at the functions of formal explanations from a different angle. Participants may just refill the missing context to interpret the question, since in real communication, the explanation process also involves two stages: the contextual interpretation of the question and the generation of the explanation (Chin-Parker & Bradner, 2017). Hence, how participants interpret questions from potential "questioners" may be one of the key factors in understanding their evaluation of formal explanations.

3. What do we know about the pragmatics of formal explanations?

Thus, while there is a substantial body of research on the pragmatics of explanations in general, much less is known about the pragmatics function of formal explanations. In what situations are they expected and appropriate? Prasada makes the following suggestion: "If someone asks why a given thing has the shape it does, providing a formal explanation makes sense if you suspect that the person does not know what kind of object the thing is" (Prasada, 2017, p. 1484), but he also adds that this hypothesis should be tested experimentally (ibid). A subsequent study on the generation of formal explanations (Haward, Wagner, Carey, & Prasada, 2018) also suggested that the form of the question (e.g., *Why are both of these yellow?*) leads participants to assume that the questioner is unaware of what category the objects belong to, as the question uses a pronoun instead of a categorical label. However, in a more recent article (Rivera et al., 2023), which claims that formal explanations have a self-sufficient explanatory function, we find the hypothesis that the questioner's awareness should not significantly affect the evaluation of the explanation's satisfactoriness: "Though it is likely true that in most cases in which a formal explanation (e.g. Mittens has whiskers because she is a

cat) is used, the questioner knows the category identity of the thing whose property is to be explained (e.g. that Mittens is a cat) and the questioner also already knows that the thing has the property whose explanation is being sought (e.g. that Mittens has whiskers), these facts do not preclude formal explanations being informative and explanatory" (Rivera et al., 2023, p. 11). Thus, the theory of formal explanations currently lacks clear answers regarding the role of pragmatic factors in the evaluation of their satisfactoriness.

Existing studies have indeed explored the connection between formal explanations and pragmatics. First of all, Vasilyeva et al. (2017) showed that utilitarian usefulness for solving a generalization task enhances persuasiveness in the case of formal explanations as well. Explanations like "Because it's a glenta, and glentas rise to the ocean's surface during the day" (ibid, p. 1438) will seem more satisfying when the judgment about the presence of the same feature in a second object is to be made based on the knowledge of its belonging or not belonging to the same category (Yes, it's a glenta / No, it's NOT a glenta // Do you think it rises to the ocean during the day?).

Other evidence indicates that people can evaluate the usefulness of formal explanations not only for themselves but also for another person. For example, Hemmatian and Sloman (2018) found that label entrenchment is an important factor that makes formal explanations more persuasive. They showed that for a formal explanation to work, the questioner must have at least the potential to understand the meaning of the categorical label, that is, it must exist within a "community of knowledge" (Sloman & Rabb, 2016) rather than being invented by the answerer. Later studies showed that the effect is driven not by how familiar the question-asking character is with the label used in the explanation, but by whether that label itself is entrenched or unentrenched within the community (Hemmatian et al., 2019). Consequently, when evaluating the persuasiveness of formal explanations, people may consider the perspective of the questioner in terms of whether it enables the explanation to be understood and utilized in the future.

Thus, although several studies have indeed touched upon certain pragmatic aspects of formal explanations, the hypothesis formulated by Prasada (2017)—namely, that formal explanations are pragmatically appropriate when the questioner is unaware of the object's category membership—has not been directly tested. To the best of our knowledge, the closest approach to testing it was taken by Gelman et al. (2018, Experiment 1). They compared whether the satisfactoriness of formal explanations changes depending on whether the question uses a pronoun or a categorical label (e.g., "Why do dogs have four legs?" vs. "Why do those [pointing to some dogs] have four legs?"). The authors found that formal explanations (e.g., "Because they are dogs") appeared less convincing if the question included a categorical label. This allows them to suggest that "it seems likely that pragmatic factors are at play" (ibid, p. 47). The authors do not dig deeply into exploring this pragmatic aspect, as the main part of their paper is devoted to testing other hypotheses. However, even this allows them to suggest that whether the questioner knows the category identity of the object may be one of the key factors for understanding formal explanations. At the same time, it should be noted that the situation described in the experimental condition with the categorical label in their study appears somewhat artificial ("Why do dogs have four legs?"—"Because they are dogs") and is unlikely to occur often in real communication, as people typically notice such obviously

circular formulations and tend to avoid them (Baum, Danovitch, & Keil, 2008; Mercier, Sudo, Castelain, Bernard, & Matsui, 2018; Rips, 2002). Therefore, although valuable conclusions follow from this experiment, its replication in a somewhat modified form might be useful.

In summary, pragmatic factors can indeed influence the persuasiveness of formal explanations. When evaluating the formal explanations, people consider: (1) to what extent the explanation will be useful to them for solving later tasks (Vasilyeva, 2017); (2) whether the questioner can use it to gain potential information within their community (Hemmatian & Sloman, 2018); and (3) whether the questioner indicates awareness of the object's category identity by directly using the categorical label in their question (Gelman et al., 2018). However, it remains unclear whether knowledge of category identity is considered every time people assess the persuasiveness of formal explanations (even in cases where the questioner does not explicitly communicate their knowledge), and some researchers argue that formal explanations may be satisfactory even when the questioner is fully aware of the object's category identity (Rivera et al., 2023). Thus, the question remains open: does the perceived awareness of the questioner regarding the object's category influence how persuasive a formal explanation seems?

4. The potential biases in understanding the pragmatics of formal explanations

Not much is known about the pragmatic factors influencing the evaluation of formal explanations, and even less can be said with certainty about the extent to which individual differences affect participants' understanding of these factors. However, we should expect a significant influence from cognitive style, as people are often vulnerable to various biases when evaluating their own and others' explanations. For instance, the so-called "illusion of explanatory depth" (Rozenblit & Keil, 2002) is well-known, where people overestimate the quality of their explanatory knowledge. Participants may also be influenced by the "seductive allure" of scientific jargon, even when it is logically irrelevant (e.g., Bennett & McLaughlin, 2024; Bulut, Gürsoy, Yorguner, Çarkaxhiu Bulut, & Sayar, 2022; Hopkins, Weisberg, & Taylor, 2016; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008, etc.). People tend to be less accurate in judging how convincing their explanations appear in terms of satisfaction, compared to accuracy (Sulik, van Paridon, & Lupyan, 2023), tend to confuse a satisfactoriness triggered by an explanation with its actual informativeness (Liquin & Lombrozo, 2022), and their preference for the order in which to receive different information in explanation conflicts with how information should be presented for more effective learning (McCarthy, Betz, & Keil, 2024).

Indirect evidence suggests that heuristics can also influence evaluations of the satisfactoriness of formal explanations. For example, a circular explanation (e.g., "*This person knows many languages because he is a person who knows many languages*") will seem more convincing if it is camouflaged as a formal explanation ("*This person knows many languages because he is a polyglot*") by replacing the circular part with an appropriate label (Aslanov & Guerra, 2023). This effect is independent of whether participants know the meaning of the label, for example, whether they know what the word polyglot means (Aslanov & Guerra, 2024). This indicates that in this case, it is not the kind of representation that is active (because for participants who do not know the meaning of the label, there is no link between the label

and the kind), but some other mechanisms. We suggest that a heuristic, intuitive evaluation is involved, reacting to the linguistic form of such explanations, that is, participants automatically recognize them as explanations similar to those that make sense in certain situations, but do not think about their content.

A similar idea was proposed by Hemmatian and Sloman (2020), applying a dual-process theory of higher cognition (see Evans & Stanovich, 2013) to describe the mechanism of formal explanations, suggesting that the initial evaluation of satisfaction with formal explanations is triggered by an intuitive system. In particular, these assumptions are based on the findings of Hemmatian and Sloman (2018), who showed that the entrenchment effect persisted even when the categories were incoherent and lacked any potential for epistemic knowledge—whether that knowledge was already shared by others in the community or could possibly be acquired in the future. Participants were influenced because they were unable to distinguish between the epistemic and social utility of the labels, which is consistent with the functioning of the intuitive system.

If this is true, we would expect that ratings of the satisfactoriness of formal explanations would depend on participants' cognitive style, that is, the extent to which they tend to give "intuitive" or "analytical" answers. There are several approaches to measuring this tendency (Newton, Feeney, & Pennycook, 2024). One widely used and effective tool for doing it is the Cognitive Reflection Test (CRT) (Frederick, 2005), which indicates which of two systems a person tends to rely on when solving problems and predicts outcomes for different heuristics-and-biases tasks (Toplak, West, & Stanovich, 2011).

5. Current studies

In six experiments, with different samples of participants, we test whether the satisfaction with a formal explanation depends on the understanding of whether the questioner is aware of the object's category identity, trying to clarify the varying assumptions made in the literature on this issue. In Experiments 1a and 1b, we test whether participants will evaluate the satisfaction with a formal explanation differently by manipulating information about whether the questioner knows the category identity. Thus, it conceptually replicates the results of Gelman et al. (2018, Experiment 1), but with a more delicate manipulation (the questioner does not use a categorical label in their question). In Experiments 2a and 2b, we explore the reverse process, that is, we provide participants with information on whether the questioner was satisfied with the formal explanation to test whether they infer from this if the questioner knew the category identity. In Experiment 3, we also manipulate information about whether the questioner was satisfied with the formal explanation, but instead of a closed-ended question, we directly asked participants why (in their opinion) the questioner asked their question. Finally, in Experiment 4, we test whether the evaluations of tautological and nontautological formal explanations differ using a set of fictional categories. Tautological formal explanations are structured such that the label refers to all objects possessing the feature that must be explained (and, therefore, cannot specify the category identity). Nontautological formal explanations use a label associated with one of several competing categories (and hence

specify the category identity). To determine whether differences in evaluations are linked to the participants' cognitive style, we used the CRT.

6. Experiments 1a and 1b

6.1. Participants

We recruited participants by distributing an advertisement among the online communities and student groups of several Russian universities. Participants were invited to voluntarily take part in a study on understanding explanations and did not receive any financial compensation. They were given the option to provide their age and gender, though this was not a mandatory requirement. However, all participants chose to share this information. Additionally, participants completed an attention check during the experiments. The attention check was the final question, which visually mirrored the structure of the stimuli (i.e., it was formatted as a dialogue). However, it actually contained an instruction to provide a specific response on a Likert scale. We excluded from the sample all participants who gave incorrect answers to this question. In Experiment 1a, all participants passed the check; in contrast, in Experiment 1b, five participants were excluded for failing it. Thus, the final sample consisted of 64 participants in Experiment 1a ($M_{age} = 30.3$ years; 42 females, 22 males) and 51 participants in Experiment 1b ($M_{age} = 30.8$ years; 37 females, 14 males). Experiment 1a was preregistered (see https://osf.io/6vzxg/; the remaining experiments in this paper were not preregistered), and according to the preregistration, we aimed to collect responses from 50 participants to calculate effect size and determine whether we needed to increase the sample size. However, since significantly more participants responded to the first call, we decided to include all 64 responses and conducted the preregistered procedures on this larger sample.

6.2. Materials and design

The materials consisted of short dialogues where characters asked their friends questions. In creating the materials, we used domains, categories, and properties taken from Rivera et al. (2023, Table 1). We developed 12 such dialogues, four for each of the three domains (animals, plants, materials). Each dialogue could appear in one of two conditions. In one condition, participants were told that the questioner was aware of the object's category identity, while in the other condition, they were told that the questioner was unaware. An example of such a dialogue in both conditions is presented below:

Peter is looking at a flying creature. Then he realizes he is looking at a bird [Condition 1] // Peter is looking at a flying creature. However, he does not realize that he is looking at a bird. [Condition 2]

He asks his friend:
- *Why does this creature fly?*
- *It flies because it's a bird.*

In Experiment 1a, a Latin square design was used to create two lists of materials so that each dialogue appeared in only one of the two conditions. Thus, the experiment followed a within-subject design. In Experiment 1b, we followed a between-subject design, and thus participants only responded to one of the two conditions, yet, there were again two experimental lists, each containing items only in one of the two conditions.

6.3. Procedure

The experiment was conducted via the Qualtrics platform, and each participant was randomly assigned to one of the two lists. Participants read dialogs presented in one of two conditions in Experiment 1a, or only in one condition in Experiment 1b. The dialogues were shown in randomized order. After each dialog, participants were asked to answer the following question: "How satisfactory do you think [character name] would find that explanation?" The questions were presented on the same screen as the dialogs. The ratings were given on a 7-point Likert scale, where 1 corresponded to "absolutely unsatisfactory" and 7 to "absolutely satisfactory."

6.4. Data analysis

We analyzed the participants' satisfaction ratings using a linear mixed-effects regression model approach, which examined the effect of the questioner's awareness of the object's category identity on the perceived satisfaction with the explanations. The fixed effect in the model was condition, which had two levels: whether the questioner was aware (aware) or unaware (unaware) of the object's category identity. To account for random variability associated with individual participants and specific items (dialogs), we included random intercepts and random slopes for condition by participants and by items in Experiment 1a, but a random slope only by item in Experiment 1b, given the design. We also conducted a post-hoc power analysis using a simulation-based approach (Brysbaert & Stevens, 2018; Kumle, Võ, & Draschkow, 2021)—see Appendix.

6.5. Results

In both experiments, the analysis revealed a significant effect of condition on participants' satisfaction ratings. Specifically, participants judged the explanations to be less satisfactory when the questioner was aware of the object's category identity compared to when the questioner was unaware of it (Experiment 1a: $\beta = -0.61$, $se = 0.24$, $t = -2.51$, $p = .015$; Experiment 1b: $\beta = -1.23$, $se = 0.37$, $t = -3.36$, $p = .001$). This indicates that the questioner's awareness of the category identity negatively influenced how satisfactory participants found the explanations (see Fig. 1).

6.6. Discussion

Experiments 1a and 1b provide evidence in favor of the hypothesis that the satisfactoriness of formal explanations can be based on their special pragmatic function, namely, their ability to indicate the category identity of an object. Also, as in Gelman et al. (2018, Experiment

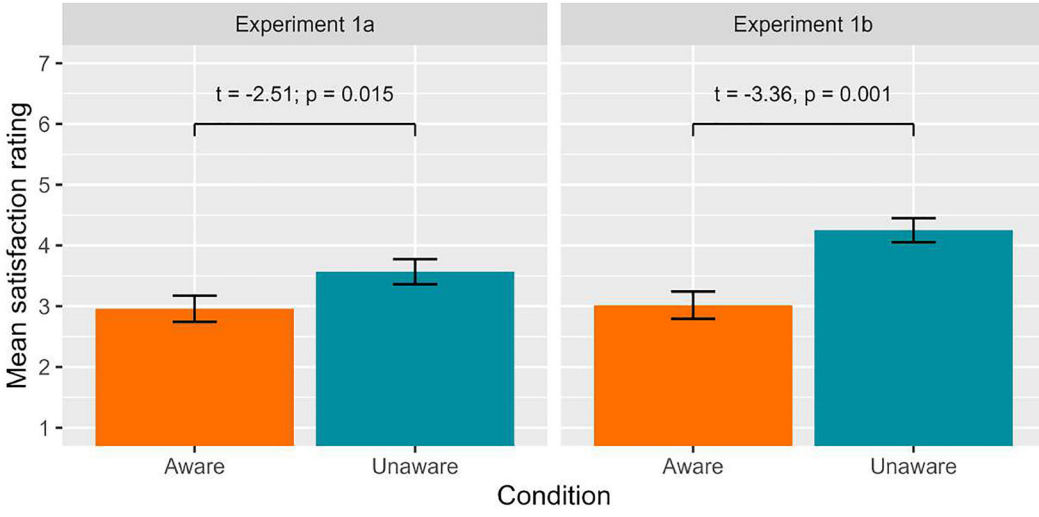

Fig. 1. Mean rating for Experiment 1a (within-subject design) and Experiment 1b (between-subject design), as a function of the experimental condition. Error bars represent 95% confidence intervals (adjusted for within-subject design for Experiment 1a).

1), participants find that formal explanations are more satisfactory when the questioner is unaware of the category identity. In this experiment, formal explanations answered the same questions and explained the same properties, and the only factor that varied between conditions was the pragmatic information about whether the questioner was aware of the object's category identity. If the persuasiveness of formal explanations were based only on their ability to represent principled connections, we would observe no difference between the two conditions. However, we see that the pragmatic factor is indeed considered by participants. Moreover, the effect is observed regardless of the level of design type (between-subjects or within-subjects). Thus, these results demonstrated that knowing the questioner is aware of categorical identity decreases formal explanations' satisfactoriness. However, in typical situations, people do not have reliable access to the questioner's knowledge and must infer it based on their responses. Therefore, in the following experiment, we used a more ecologically valid design. If the persuasiveness of formal explanations depends precisely on their ability to signal categorical identity, participants would need to conclude that, in cases where the formal explanation was satisfactory, the questioner was unaware of the object's categorical identity.

7. Experiments 2a and 2b

7.1. Participants

We used the same recruitment strategy as in Experiments 1a and 1b. The attention check (identical to the one described for Experiments 1a and 1b) led to the exclusion of two participants from Experiment 2a and five participants from Experiment 2b. The final sample

included 80 participants ($M_{age} = 27.1$ years; 56 females, 24 males) in Experiment 2a, and 55 participants ($M_{age} = 31.3$ years; 40 females, 15 males) in Experiment 2b. They did not receive financial rewards for their participation.

7.2. Materials, design, procedure, and data analysis

We used the same materials as in Experiments 1a and 1b, but with slight modifications. This time, we did not report whether a questioner knew the category identity. Instead, after each dialog, we reported whether a questioner was satisfied with the explanation. Consequently, after each dialog, participants had to guess whether the character knew about the category identity. To do this, they answered the corresponding question on a 7-point scale, for example: *Do you think [character name] was more likely to realize or not realize that it was a bird when he asked his question*? In this question, 1 corresponded to "absolutely did not realize" and 7 corresponded to "absolutely did realize."

Peter looks at a flying creature.

He asks his friend:
- *Why does this creature fly?*
- *It flies because it's a bird.*

Peter was satisfied with this explanation [Condition 1] // Peter was not satisfied with this explanation [Condition 2]

Latin square design was applied, so each dialog appeared only once and only in one of the two conditions in Experiment 2a. Thus, two lists of items were designed, and participants were randomly assigned to one of them. In Experiment 2b, we used a between-subjects design, so we also constructed two experimental lists; however, in this case, each list corresponded to only one experimental condition.

7.3. Results

The analysis of Experiments 2a and 2b also revealed a significant effect of condition on participants' ratings. Specifically, participants judged the questioner to be less aware of the categorical identity of the object, when the questioner was presented as satisfied with the explanation provided in Experiment 2a ($\beta = -2.18$, $se = 0.25$, $t = -8.82$, $p < .001$) and Experiment 2b ($\beta = -0.97$, $se = 0.40$, $t = -2.46$, $p = .017$). This indicates that the questioner's satisfaction regarding the explanation provided is highly dependent on their awareness of category identity, that is, participants believe that if formal explanations are not satisfactory, then the questioners know what category the object belongs to (see Fig. 2).

7.4. Discussion

Experiments 2a and 2b show that when the questioner is satisfied with a formal explanation, people assume that the questioner was unaware of the categorical identity of the object. Together with Experiments 1a and 1b results, this provides solid grounds to suppose that

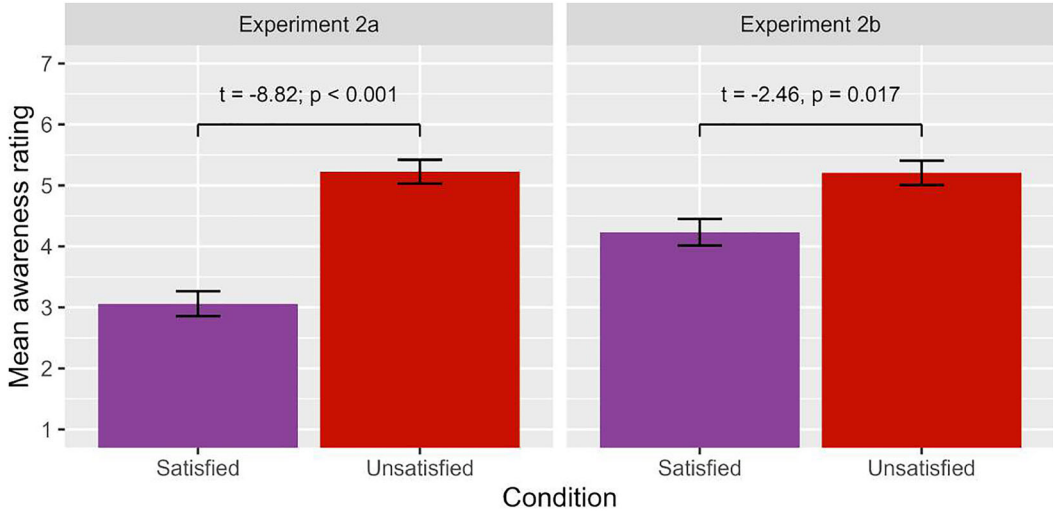

Fig. 2. Mean rating for Experiment 2a (within-subject design) and Experiment 2b (between-subject design), as a function of the experimental condition. Error bars represent 95% confidence intervals (adjusted for within-subject designs for Experiment 2a).

satisfaction with such explanations significantly depends on a pragmatic factor. It appears that when participants evaluate formal explanations in experimental settings, they reconstruct the typical circumstances in which such explanations are used and try to imagine a situation where simply pointing out the categorical identity would have explanatory power. The most obvious scenario here is that the questioner does not know the categorical identity.

This is why participants thought the questioner would be less satisfied when aware of the categorical identity (Experiments 1a and 1b). Similarly, if a formal explanation is found satisfactory, participants assume that the inquirer was not aware of the category (Experiments 2a and 2b). However, in previous experiments, we somewhat framed participants, leading them to see a connection between the satisfaction with explanations and knowledge of categorical identity, as we directly asked them to assume the probability of such a connection. It is possible that in other circumstances, participants might not have considered this connection or identified it as a key condition for the satisfactoriness of formal explanations. Therefore, in Experiment 3, we chose an open-ended response approach so that participants would not be restricted in how they interpret the pragmatic functions of formal explanations.

8. Experiment 3

8.1. Participants

Using the same recruiting strategy, we invited 35 participants. However, two of them did not provide full responses, so they were excluded from the final sample. Thus, we analyzed replies from 33 participants ($M_{age} = 26.4$ years; 20 females, 13 males).

8.2. Materials and design

We used the same set of materials as in Experiment 2. However, since this time the question was open-ended, that is, more time-consuming to answer, we reduced the number of dialogs, leaving only two dialogs in each of the three domains. The design was also the same, that is, each dialog was presented in one of the two conditions (the characters were satisfied with the explanation vs. they were not satisfied), and participants were randomly assigned to one of two experimental lists.

8.3. Procedure

We followed the procedure described in Experiment 2. However, instead of presenting a closed-ended question, we asked the participants to answer an open-ended question (e.g., *Why do you think Peter had this question?*). We did not limit responses in terms of either the characters or the number of guesses.

8.4. Data analysis

We conducted a content analysis of participants' responses and grouped them into six possible types of answers. Since each response could contain multiple guesses, a participant's response could belong to more than one type of answer. The identified six types of responses to which participants' guesses belong are summarized in Table 1.

After the content analysis, participants' responses for each item were given a 1 under each of the types of answers that were contained in their responses. Whenever the response did not include a given type of answer, this one received a 0. Consequently, participants' responses were transformed from an open question to a binomial distribution, with which we calculated the percentage of occurrences for each category under the two levels of the experimental condition (character satisfied vs. character not satisfied). Inferential analysis was carried out using a generalized linear mixed-effect binomial regression.

8.5. Results

The results of the generalized mixed-effect regression are shown in Table 2. As it can be seen (Fig. 3), the Category type of answer exhibits a large difference in the probability of appearing as a function of the experimental condition, as reflected by the significant main effect of condition ($\beta = 2.15$, $se = 0.36$, $z = 5.94$, $p < .001$). We also observed a significant main effect of answer type when comparing the overall probability of Category and Function to appear as answer ($\beta = -2.55$, $se = 1.29$, $z = -1.98$, $p = .048$). More importantly, Category had the highest probability of appearing as an answer (compared to the other five types of answers) when the characters were satisfied, and the lowest probability of appearing as an answer (compared to the other five types of answers) when the characters were not satisfied. This is reflected by the reliable interaction effects.

8.6. Discussion

Experiment 3 provides clear evidence that the satisfactoriness of formal explanations is related to their ability to indicate the category identity of an object. Participants tended to

Table 1
Categories and correspondent examples of participants' responses in Experiment 3

Type of answer	Examples	Definition
Categorical	Peter probably did not know that this creature could be a bird (e.g., it does not look like a bird) Nicholas wondered exactly what kind of metal this shiny metal was	The characters asked questions because they could not determine which category the object belonged to
Mechanistic	He is interested in what substances provide the color of the fruit Perhaps she's interested in a deeper explanation in terms of the physics of magnetism	The characters were interested in the biological, physical, anatomical, and so on mechanisms that cause the feature to exist
Functional	Elena apparently knows they are roses. She probably asked why thorns grow on flowers like this in general, if they have some kind of function Perhaps he knows that the shine of metal depends on the treatment and he wonders why the metal was polished	The characters were interested in the purpose of the feature or the possible benefit of it
Causal	Because she wanted to know what made those thorns appear Svetlana wanted to know the reasons why evolutionarily zebras became striped	The characters were interested in what caused the feature (but it is impossible to define their interest more precisely)
Personal	Elena ordered roses with already cut thorns, and received regular ones Curiosity, a desire to learn about the world	Some particular circumstance or personal trait caused the character to ask a question
Other	Svetlana doesn't understand why zebras have striped colors and decides to ask her friend, but he gives an uninformative answer He wonders why, on principle, anyone flies	Generalized or unspecific answers

assume that if the questioners were satisfied with a formal explanation, they wanted to learn about the categorical identity of the object. At the same time, if the questioners were dissatisfied with the formal explanation, their question was motivated by a variety of possible reasons, excluding the desire to learn about the category identity. Thus, although the experimental materials lack any pragmatic information except whether the questioner liked the explanation, participants reconstruct the communicative situation unambiguously. From their point of view, formal explanations are convincing when the questioner recognizes a feature of a category but cannot identify the category itself. Taking together the results from these three experiments, we can assume that formal explanations fulfill a pragmatic pointing function, since they are satisfactory only when the questioner needs an indication of categorical identity.

Table 2
Results of the generalized linear mixed-effects regression model

	β	*se*	*z*	*p*	*Sig.*
(Intercept)	−1.96	0.50	−3.95	< .001	***
Condition	2.15	0.36	5.94	< .001	***
Category versus Cause	−1.11	0.73	−1.53	.127	
Category versus Function	−2.55	1.29	−1.98	.048	*
Category versus Mechanism	−0.39	1.06	−0.37	.713	
Category versus Person	−0.52	0.79	−0.66	.509	
Category versus Other	0.01	0.59	0.02	.985	
Condition * Category versus Cause	−2.95	0.48	−6.12	< .001	***
Condition * Category versus Function	−4.01	0.67	−6.01	< .001	***
Condition * Category versus Mechanism	−3.40	0.47	−7.30	< .001	***
Condition * Category versus Person	−2.43	0.44	−5.58	< .001	***
Condition * Category versus Other	−2.02	0.42	−4.85	< .001	***

*** = $p < .001$; * = $p < .05$.

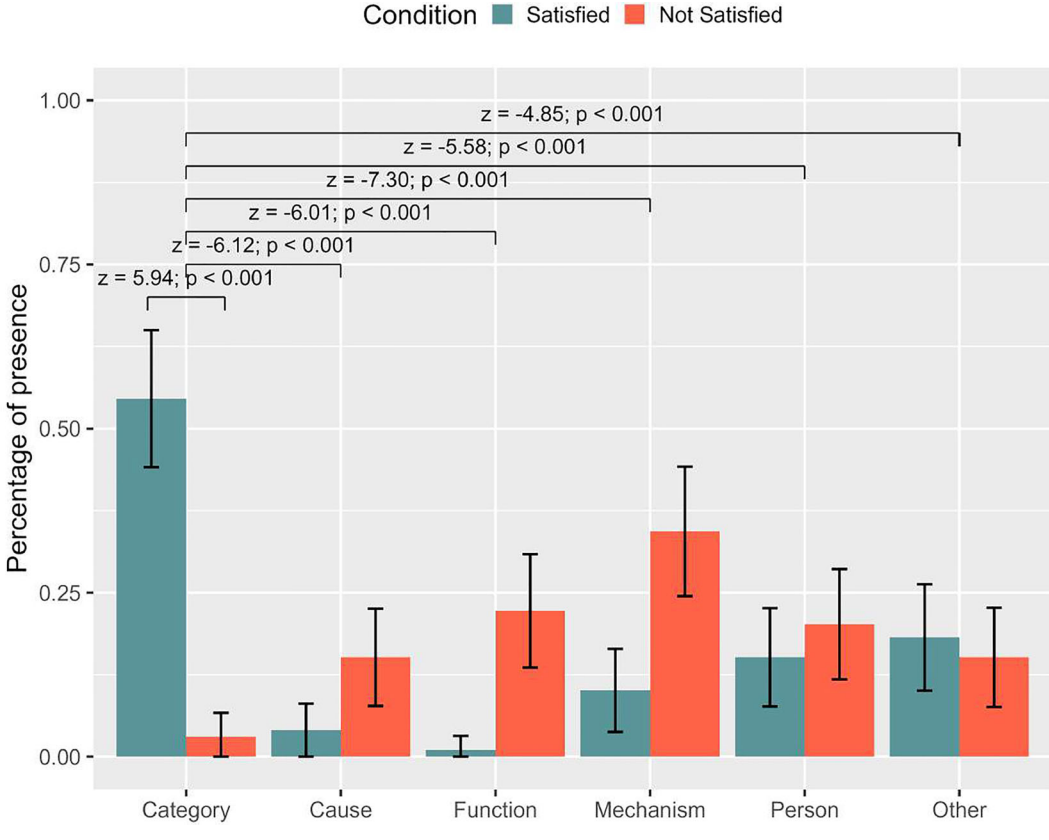

Fig. 3. Percentage of the presence of each category as a function of condition. Error bars represent 95% confidence intervals. Critical contrasts (main effect of condition on the Category type of answer, and all interactions) are presented visually with *t*- and *p*-values from the binomial regression.

To test this claim, in Experiment 4, we directly compared formal explanations that can fulfill such a function and formal explanations that are implicitly circular and lack such a function since they cannot point to any specific category. In this regard, given that the evaluation of formal explanations could potentially be associated with some heuristics (Hemmatian & Sloman, 2020), such as focusing on the formal structure of the explanation without considering its content (Aslanov & Guerra, 2023; Aslanov & Guerra, 2024), we added cognitive reflection as a factor influencing perceptions of the pragmatic relevance of formal explanations.

9. Experiment 4

9.1. Participants

We used the same recruitment strategy as in previous experiments. We collected responses from 138 participants, but six were deleted since those participants failed the attention check (identical to the one described earlier). Thus, the final sample consisted of 132 participants ($M_{age} = 30.8$; 92 females, 39 males, 1 preferred not to specify their gender).

9.2. Materials and design

We designed a set of vignettes, each containing either a formal or a circular explanation. Participants were asked to imagine that they were researchers who studied an alien civilization and had arrived on a different planet to study how people explained things to each other there. Each vignette was preceded by an instruction that informed participants that they had learned a new word from an alien language (a pseudoword). These instructions were present on the same screen as the explanations, so that participants did not need to memorize pseudowords. Depending on the experimental condition, a pseudoword meant either *all objects* with a feature to be explained or *one of three possible categories*, each of which had that feature. When a pseudoword meant all objects with a feature to be explained, its definition literally repeated the explanandum, creating an implicitly circular explanation. For example, for the explanation "*Why does this animal have hooves? - Because it is a litus*," the definition of the pseudoword "*litus*" was "*an animal that has hooves*" in that condition. In the category-choice condition, participants were told that there were only three types of hooved animals on that planet: one was called "*styro*," another "*satu*," and the third "*litus*." Thus, we constructed materials for four experimental conditions: "tautology" (a pseudoword means all objects + formal explanation), "choice" (a pseudoword means one of three categories + formal explanation), "control 1" (a pseudoword means all objects + circular explanation), and "control 2" (a pseudoword means one of three categories + circular explanation). Both control conditions contained one of the two instructions from the experimental conditions, but their explanations were explicitly circular and did not use pseudowords (see Table 3). It is important to note that, whereas in the previous experiments, we asked participants to evaluate the explanations from the perspective of the questioners, in the present experiment, there was no such instruction. What was important to us was that participants focused on whether the explanation could (or could not) indicate a specific category, thereby serving the pragmatic function of specifying a category.

Table 3
Example of creating conditions from two types of instructions and two types of explanations

Example of explanations \ Example of instructions	You learned that their language includes the word "*kirto*" which means "a plant with thorns"	You learned that there are only three kinds of plants with thorns on their planet. One is called "tivo," one is called "tala," and the third is called "*kirto*"
You observe two aliens communicating with each other. One of them points toward some mysterious plant with thorns. He then asks his friend: Why does this plant have thorns? His friend replies: Because it's a *kirto*	"Tautology" condition	"Choice" condition
You observe two aliens communicating with each other. One of them points in the direction of some mysterious plant with thorns. He then asks his friend: Why does this plant have thorns? His friend replies: Because it's *a plant with thorns*	"Control 1" condition	"Control 2" condition

We designed 16 explanations, four each in the animal, plant, mineral, and artifact (jewelry) domains. We created four experimental lists of items using a Latin square design so each explanation appeared only once and only in one of the conditions in each of the lists.

9.3. Procedure

The experiment was conducted via the Qualtrics platform. Participants were randomly assigned to one of four experimental lists. The dialogues were shown in randomized order. After reading each dialog, they had to answer the question "How satisfactory do you find this explanation?". To do so, they gave a mark on a scale from 1 to 7, where 1 corresponded to "absolutely unsatisfactory" and 7 to "absolutely satisfactory." After the experimental part, they answered the CRT questionnaire (Frederick, 2005), which has been adapted for the Russian language (Rodina & Prudkov, 2019).

9.4. Data analysis

Participants' satisfaction ratings were analyzed using a linear mixed-effects model to assess the effects of explanation condition and CR ability on the perceived satisfactoriness of explanations. The fixed effects included condition (with four levels: tautology, choice, control 1, and control 2), CRT score, and their interaction. Random intercepts and slopes for condition were included for participants and items to account for individual differences and item variability. The items random intercept also included a random slope of CR. The main model had as reference group the Tautology level and the lowest value of CR. We rotated the intercept to verify that the interaction effect of interest was present exclusively for the interaction between CR and Tautology-Choice differences, and not across conditions.

In addition to the primary mixed-effects model, supplementary analyses were performed to elucidate the nature of the interaction between condition and CR. First, the simple slopes of CR were estimated within each level of the categorical predictor. To do this, we computed the conditional effect of CR on the satisfaction rating separately for each experimental condition. This step allowed for an assessment of whether the magnitude and direction of the CR effect varied systematically across conditions. Subsequently, pairwise comparisons among the levels of condition were conducted to test the differences in the CR levels. Contrasts were calculated at each of the four CR values to determine if the observed differences between conditions were significant. The analysis produced estimates for each contrast, z-statistics, p-values, and 95% confidence intervals. These comparisons provided a more granular interpretation of the interaction by identifying which pairs of conditions exhibited differential relationships between CR and the rating, while maintaining the overall structure and interpretation of the primary model.

9.5. Results

Table 4 shows the results of the linear mixed-effects regression analysis. The analysis revealed two significant main effects of condition on participants' satisfaction ratings. Specifically, participants rated explanations in the control conditions as significantly less satisfac-

Table 4
Results of the linear mixed-effect regression model

	β	*se*	*t*	*p*	*Sig.*
(Intercept)	3.54	0.30	11.86	<.001	
Tautology versus Choice	−0.26	0.24	−1.05	.298	
Tautology versus Control 1	−1.17	0.28	−4.26	<.001	***
Tautology versus Control 2	−1.26	0.29	−4.36	<.001	***
CR	−0.16	0.14	−1.10	.272	
Tautology versus Choice * CR	0.24	0.11	2.09	.039	*
Tautology versus Control 1 * CR	0.08	0.13	0.60	.548	
Tautology versus Control 2 * CR	0.06	0.14	0.41	.682	

*** = $p < .001$; * = $p < .05$.

Table 5
Estimated slopes of CR by condition

Slope	Estimate	*se*	*z*	*p*
Tautology	−0.16	0.14	−1.10	.273
Choice	0.08	0.14	0.57	.572
Control1	−0.08	0.11	−0.70	.487
Control2	−0.10	0.11	−0.94	.348

tory compared to the Tautology condition. For the Control 1 condition, the satisfaction ratings were lower ($\beta = -1.17$, $se = 0.28$, $t = -4.26$, $p < .001$), and similarly for the Control 2 condition ($\beta = -1.26$, $se = 0.29$, $t = -4.36$, $p < .001$). There was no significant difference between the choice condition and the tautology condition. The CR scores showed the following distribution: 27 participants with CR = 0; 27 participants with CR = 1; 32 participants with CR = 2; and 46 participants with CR = 3. A significant interaction between condition (i.e., Tautology vs. Choice) and CR was observed ($\beta = 0.24$, $se = 0.11$, $t = 2.09$, $p = .039$). No other main or interaction effects were found. Finally, when we rotated the intercept to see whether CR interacted with the difference between Tautology and the Control conditions, we found no reliable interactions (both *t*-values < 1), reflecting that CR distinctively modulates the relationship between Tautology and Choice.

To further clarify the interaction effect, we estimated the simple slopes of CR for each condition using the *marginaleffects* package for R (Arel-Bundock, Greifer, & Heiss, 2024) (Table 5). In the Tautology condition, the slope for CR was estimated at –0.16 ($z = -1.10$, $p = .273$). In contrast, the slope in the Choice condition was 0.08 ($z = 0.57$, $p = .572$), while the slopes for the Control 1 and Control 2 conditions were –0.08 ($z = -0.70$, $p = .487$) and –0.10 ($z = -0.94$, $p = .348$), respectively. These estimates suggest that the effect of CR on satisfaction ratings varies in both magnitude and direction across the experimental conditions, with the notable divergence between the Tautology (most negative) and Choice (most positive) conditions. Finally, pairwise contrasts were then computed at fixed CR values (0, 1, 2, and 3) to assess differences between the Tautology condition and each of the other conditions.

Table 6
Results of the comparisons of marginal effects of condition by CR

Contrast	*CR*	β	*se*	*z*	*p*	*Sig.*
Choice versus Tautology	0	−0.12	0.24	−0.49	.624	
Control1 versus Tautology	0	−1.16	0.28	−4.22	<.001	***
Control2 versus Tautology	0	−1.24	0.29	−4.31	<.001	***
Choice versus Tautology	1	−0.05	0.16	−0.32	.749	
Control1 versus Tautology	1	−0.80	0.18	−4.34	<.001	***
Control2 versus Tautology	1	−0.88	0.19	−4.63	<.001	***
Choice versus Tautology	2	−0.07	0.14	−0.53	.593	
Control1 versus Tautology	2	−1.54	0.16	−9.41	<.001	***
Control2 versus Tautology	2	−1.73	0.16	−10.50	<.001	***
Choice versus Tautology	3	0.61	0.20	3.13	.002	**
Control1 versus Tautology	3	−0.74	0.24	−3.15	.002	**
Control2 versus Tautology	3	−0.88	0.24	−3.67	<.001	***

*** = $p < .001$; ** = $p < .01$.

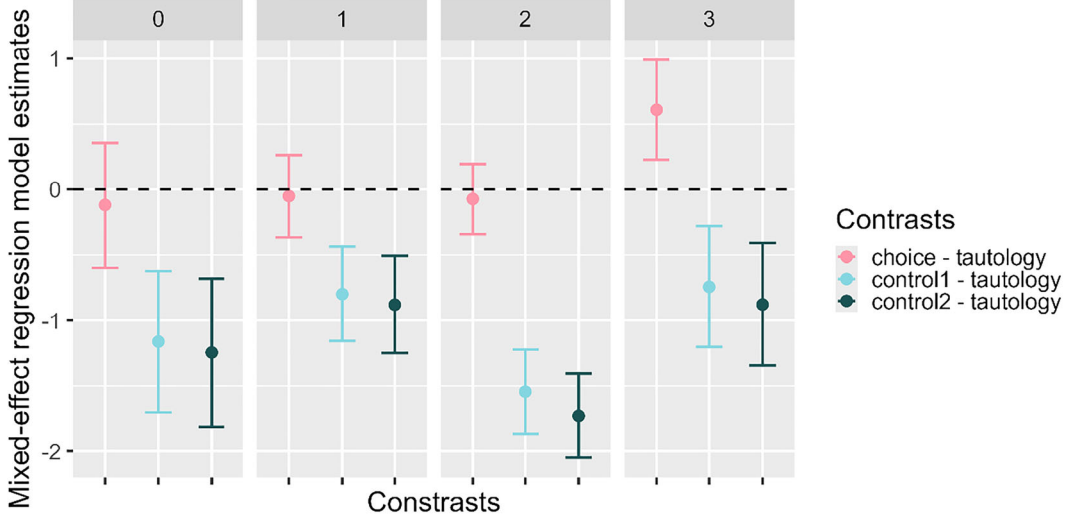

Fig. 4. Pairwise differences in predicted satisfaction ratings from the linear mixed-effects model between each condition (Choice, Control 1, and Control 2) and the reference condition (Tautology) at four levels of CR (0, 1, 2, and 3). Each point represents the estimated contrast in satisfaction ratings, with vertical bars indicating the 95% confidence intervals. The dashed horizontal line at 0 marks no difference from Tautology. Pink circles depict the Choice–Tautology contrast, while light and dark blue circles show the Control 1–Tautology and Control 2–Tautology contrasts, respectively.

The results are presented in Table 6 and Fig. 4. The difference between Tautology and the two control conditions, across CR values, is statistically significant (all p-values $< .01$). Importantly, the difference between Tautology and the Choice conditions was significant only at the highest level of CR ($\beta = 0.61$, $se = 0.20$, $z = 3.13$, $p = .002$).

9.6. Discussion

In Experiment 4, we tested the hypothesis that the satisfactoriness of formal explanations is related to their ability to point to the categorical identity of an object. We compared tautological formal explanations (in which the label points to all objects possessing the given feature) and nontautological formal explanations (in which the label points to one of three categories, each of which possesses the given feature). As we can see, tautological formal explanations (the "tautology" condition) are perceived as more satisfactory than explicitly circular explanations (the "control 1" and "control 2" conditions), but do not differ from nontautological formal explanations (the "choice" condition). However, cognitive reflection has a significant effect on the evaluation of the satisfactoriness of formal explanations. The difference between the ratings of tautological and nontautological formal explanations emerged only at the highest level of cognitive reflection, that is, when participants showed a strong preference for analytical over intuitive reasoning.

10. General discussion

Research on formal explanations has primarily focused on their ability to establish principled connections between an instance and a kind (e.g., Prasada, 2017; Haward, Carey, & Prasada, 2021). Another line of work has explored what we call the "secondary benefits" of formal explanations—their ability to indicate potential sources of additional knowledge without directly providing it. For example, they can point to knowledge potentially existing within a community (Hemmatian & Sloman, 2018), to an underlying cause of a feature without specifying the cause itself (Gelman et al., 2018), or to a domain of prior knowledge where relevant information can be found (Aslanov et al., 2022). However, the pragmatic factors that make formal explanations more satisfactory, especially concerning the questioner's categorical knowledge, remain underexplored. What motivates questioners to seek a formal explanation, and why do they find it more or less satisfactory? To what extent does the questioner's awareness of the category identity of an object affect their evaluation of formal explanations? Some studies suggest that such knowledge is a key factor (Gelman et al., 2018, Experiment 1), while others propose that it should not significantly impact the explanatory value of formal explanations (Rivera et al., 2023).

In six experiments, we investigated the relationship between categorical knowledge and the perceived satisfaction with formal explanations. In Experiments 1a and 1b, we demonstrated that participants tend to believe that knowing the category identity makes formal explanations less satisfactory. This relationship was also distinct in Experiments 2a and 2b, where participants assumed that if the questioners were satisfied with a formal explanation, it inferred that they had not known the object's category identity. Finally, in Experiment 3, participants were asked to answer open-ended questions and provide explanations for why satisfied and dissatisfied questioners asked their questions. Again, the same trend emerged: when characters in the stimuli were satisfied with the formal explanation, participants commonly inferred

that the questioners wanted to learn the object's category. However, when the characters were dissatisfied, the desire to learn the object's category was the least frequent guess.

These experiments collectively suggest that **formal explanations serve a specific pragmatic function: they indicate the category of an object when only its features are known**. In all five experiments, participants interpreted the dialogues in a similar way, assuming that satisfied questioners asked their questions because they could not identify the object. This aligns with the results of Gelman et al. (2018, Experiment 1), where participants found formal explanations less satisfactory when the questioner used a categorical label in their question. In such cases, participants considered the formal explanation inappropriate since the questioner already knew the object's category. At the same time, our findings suggest that the satisfaction with formal explanations observed by Rivera et al. (2023) was not solely due to their ability to support kind representations. Using the same categories and features as Rivera et al. (2023), we showed a link between satisfaction and category knowledge, suggesting that participants in Rivera's experiments might have reconstructed the pragmatic context based on their own experiences, assuming that questioners were not aware of the category identity when posing their questions. Indeed, when presented with a dialogue such as "*Why does that (pointing to a tiger) have stripes? – Because it is a tiger*," it is natural to assume that the questioner simply did not recognize the tiger. Otherwise, as indicated by our Experiments 1–3, such an explanation would have been less satisfactory.

If formal explanations are perceived as more satisfying only when the questioner seeks to determine the category, it follows that explanations incapable of specifying the category would be unsatisfactory. For example, explanations like "*This person knows many languages because they are a polyglot*" should be unsatisfactory since they do not help identify the category—anyone who knows many languages is by definition a polyglot. However, it is known that such tautological formal explanations are more satisfactory than explicit circular explanations (Aslanov & Guerra, 2023; Aslanov & Guerra, 2024). This contradiction can be resolved by adopting a dual-system approach to interpreting formal explanations (Hemmatian & Sloman, 2020). As shown in Experiment 4, the ability to resist automatic responses of the intuitive system helps individuals better understand the contexts in which formal explanations are appropriate. Individuals with the highest cognitive reflection tend to find tautological formal explanations less satisfactory than nontautological ones. Thus, while formal explanations are more satisfactory when they are capable of specifying a category when it is unclear, the intuitive system can still lead us to trust these explanations even when they are pragmatically useless. One possible mechanism is that individuals with low cognitive reflection respond primarily to the form of formal explanations, as they appear natural and often receive high ratings for their natural sound (Gelman et al., 2018; Prasada & Dillingham, 2006; Rivera et al., 2023). This raises the question of how cognitive reflection influences mindreading and the ability to represent another's knowledge (see, e.g., Spaulding, 2020). Thus, the role of mindreading in understanding the pragmatic context of formal explanations is an interesting direction for future research.

Whereas previous studies (Hemmatian & Sloman, 2018) did not find an interaction between cognitive reflection and the pragmatic factor—namely, label entrenchment—Experiment 4 revealed such an interaction. This suggests that cognitive reflection does not play a universal

role in the evaluation of formal explanations, and that different types of pragmatic factors may interact with cognitive reflection in distinct ways—or not interact at all. In light of this, the results reported here should not be generalized to other pragmatic factors, and the role of cognitive reflection should be tested separately in each case.

The data from all six experiments deepen our understanding of why formal explanations may sometimes appear satisfactory. It seems that their ability to point to a specific category when only a feature is known contributes as distinctly as their role in supporting kind representations. Future research could explore which of these aspects carries more weight. This could be achieved by comparing formal explanations with principled properties (e.g., "*Fido has four legs because he is a dog*") when the category identity is known to the questioner, against formal explanations without principled properties (e.g., "*Fido has a collar because he is a dog*") when the category identity is unknown.

At the same time, a closer examination of the effect sizes in Experiments 1 and 2 reveals that the effect size differs across the between-subjects and within-subjects versions. This suggests that satisfaction with formal explanations may substantially depend both on the question asked and on whether participants have the opportunity to compare different pragmatic contexts. We did not formulate any prior hypotheses regarding this difference and, therefore, cannot offer fully satisfactory post-hoc interpretations. However, we believe that future research on formal explanations could take this aspect into account in their design, as some effects may emerge only at the within-subject or only at the between-subject level.

Moreover, our findings complement existing evidence on the role of pragmatics in evaluating explanations. It is known that individuals consider the utility of various explanations for their tasks and future information-seeking efforts (Meyer & Keil, 2024; Vasilyeva et al., 2017). Our results show that people also tend to hypothesize about the epistemic needs of questioners and try to predict which information will be most relevant based on what the questioner already knows. As demonstrated by Joo et al. (2022), participants often interpret the question "Why…?" as "What is the purpose…?" when evaluating teleological explanations. Our data suggest that when evaluating formal explanations, participants interpret the question "Why…?" as "What exactly is it?" Future research could further investigate these implicit questions, with one possible approach being to ask participants to rephrase the question in their own words.

We should also acknowledge some limitations of the present study. While in Experiments 1–3 participants were explicitly instructed to adopt the perspective of a questioner, this was not the case in Experiment 4. Although we do not believe this had a substantial impact on the results, replicating this experiment with an explicit focus on the questioner's perspective could be informative. Additionally, since we used the original three-item CRT (Frederick, 2005), its limited sensitivity may have constrained our ability to detect relevant effects. Using extended versions of the test with more items (e.g., Toplak, West, & Stanovich, 2014) might provide a more nuanced picture of how the two systems of thinking relate to the evaluation of the pragmatic functions of formal explanations. The inclusion of an additional manipulation check could also provide a more precise account of the role of cognitive reflection by better distinguishing between merely less attentive participants and those with low levels of cognitive reflection. Finally, we would like to note that only the first experiment in this series

was preregistered. Therefore, replication of the reported findings in future studies may be necessary.

Our findings emphasize that people rarely evaluate the abstract correctness of explanations; instead, they often try to understand for whom and in what context the explanation might be satisfactory. This underscores the need to consider the pragmatic context and participants' interpretations when studying the satisfactoriness of different types of explanations.

Acknowledgments

Funding from ANID Support 2024 AFB240004 is gratefully acknowledged (Ivan Aslanov, Ernesto Guerra). This work is also the output of an individual research project of Tatyana Kotova, Assistant Professor, Department of Psychology, RANEPA. Work of Alexey Kotov is supported by the Basic Research Program at the HSE University (Experiments 2a,b).

Data availability statement

Data, materials, and R scripts are openly available at the project's OSF page (https://osf.io/gbpfs/).

References

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginal effects for R and Python. *Journal of Statistical Software*, *111*(9), 1–32. doi: 10.18637/jss.v111.i09

Aslanov, I. A., Sudorgina, Y. V., & Kotov, A. A. (2022). The explanatory effect of a label: Its influence on a category persists even if we forget the label. *Frontiers in Psychology*, *12*, 745586.

Aslanov, I., & Guerra, E. (2023). Tautological formal explanations: Does prior knowledge affect their satisfiability? *Frontiers in Psychology*, *14*, 1258985.

Aslanov, I., & Guerra, E. (2024). Tautological formal explanations are satisfactory regardless of prior knowledge. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. *46*).

Banerjee, K., & Bloom, P. (2015). "Everything happens for a reason": Children's beliefs about purpose in life events. *Child Development*, *86*, 503–518.

Baum, L. A., Danovitch, J. H., & Keil, F. C. (2008). Children's sensitivity to circular explanations. *Journal of Experimental Child Psychology*, *100*(2), 146–155.

Bennett, E. M., & McLaughlin, P. J. (2024). Neuroscience explanations really do satisfy: A systematic review and meta-analysis of the seductive allure of neuroscience. *Public Understanding of Science*, *33*(3), 290–307.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20.

Bulut, N. S., Gürsoy, S. C., Yorguner, N., Çarkaxhiu Bulut, G., & Sayar, K. (2022). The seductive allure effect extends from neuroscientific to psychoanalytic explanations among Turkish medical students: Preliminary implications of biased scientific reasoning within the context of medical and psychiatric training. *Thinking & Reasoning*, *28*(4), 625–644.

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*, 227–249.

Chin-Parker, S., & Bradner, A. (2017). A contrastive account of explanation generation. *Psychonomic Bulletin & Review*, *24*, 1387–1397.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Faye, J. (2007). The pragmatic-rhetorical theory of explanation. In Johannes Persson & Petri Ylikoski (Eds.), *Rethinking explanation* (pp. 43–68). Dordrecht: Springer Netherlands.

Foster-Hanson, E., & Lombrozo, T. (2022). How "is" shapes "ought" for folk-biological concepts. *Cognitive Psychology*, *139*, 101507.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409.

Gelman, S. A., Cimpian, A., & Roberts, S. O. (2018). How deep do we dig? Formal explanations as placeholders for inherent explanations. *Cognitive Psychology*, *106*, 43–59.

Giffin, C., Wilkenfeld, D., & Lombrozo, T. (2017). The explanatory effect of a label: Explanations with named categories are more satisfying. *Cognition*, *168*, 357–369. https://doi.org/10.1016/j.cognition.2017.07.011

Haward, P., Carey, S., & Prasada, S. (2021). The formal structure of kind representations. *Cognitive Science*, *45*(10), e13040.

Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition*, *176*, 255–268.

Hemmatian, B., Chan, S. Y., & Sloman, S. A. (2019). What gives a diagnostic label value? Common use over informativeness. https://doi.org/10.31234/osf.io/er6ku

Hemmatian, B., & Sloman, S. A. (2018). Community appeal: Explanation without information. *Journal of Experimental Psychology: General*, *147*, 1677–1712. https://doi.org/10.1037/xge0000478

Hemmatian, B., & Sloman, S. A. (2020). Two systems for thinking with a community: Outsourcing versus collaboration. In S. Elqayam, I. Douven, J. S. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to David Over* (pp. 102–115). New York: Routledge.

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, *155*, 67–76.

Joo, S., Yousif, S. R., & Keil, F. C. (2022). Understanding "Why:" How implicit questions shape explanation preferences. *Cognitive Science*, *46*(2), e13091. https://doi.org/10.1111/cogs.13091

Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, *35*, 1440–1452.

Kelemen, D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological Science*, *15*, 295–301.

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543.

Liquin, E. G., & Lombrozo, T. (2022). Motivated to learn: An account of explanatory satisfaction. *Cognitive Psychology*, *132*, 101453. https://doi.org/10.1016/j.cogpsych.2021.101453

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204. https://doi.org/10.1016/j.cognition.2004.12.009

McCarthy, A. M., Betz, N., & Keil, F. C. (2024). Adults' learning of complex explanations violates their intuitions about optimal explanatory order. *Cognition*, *246*, 105767.

Mercier, H., Sudo, M., Castelain, T., Bernard, S., & Matsui, T. (2018). Japanese preschoolers' evaluation of circular and non-circular arguments. *European Journal of Developmental Psychology*, *15*(5), 493–505.

Meyer, M., & Keil, F. (2024). Circular but suggestive: Pragmatic insights from reductive tautologies. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. *46*).

Newman, G. E., & Knobe, J. (2019). The essence of essentialism. *Mind & Language*, *34*(5), 585–605.

Newton, C., Feeney, J., & Pennycook, G. (2024). On the disposition to think analytically: Four distinct intuitive-analytic thinking styles. *Personality and Social Psychology Bulletin*, *50*(6), 906–923.

Prasada, S. (2017). The scope of formal explanation. *Psychonomic bulletin & review*, *24*(5), 1478–1487.

Prasada, S. (2021). Formal explanation and mechanisms of conceptual representation. In Ludger Jansen, & Petter Sandstad (Eds.), *Neo-Aristotelian perspectives on formal causation* (pp. 269–286).
Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*(1), 73–112.
Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, *33*(3), 401–448.
Rips, L. J. (2002). Circular reasoning. *Cognitive Science*, *26*(6), 767–795.
Rivera, S., Prasad, S., & Prasada, S. (2023). Are formal explanations mere placeholders or pointers? *Cognition*, *235*, 105407.
Rodina, O. N., & Prudkov, P. N. (2019). Testing the Russian versions of the Cognitive Reflection Test. *Voprosy Psikhologii* (4), 155–162.
Rose, D. (2020). Mentalizing objects. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. *4*). Oxford: Oxford University Press.
Rose, D., & Schaffer, J. (2017). Folk mereology is teleological. *Noûs*, *51*, 238–270.
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562.
Schachner, A., Zhu, L., Li, J., & Kelemen, D. (2017). Is the bias for function-based explanations culturally universal? Children from China endorse teleological explanations of natural phenomena. *Journal of Experimental Child Psychology*, *157*, 29–48. https://doi.org/10.1016/j.jecp.2016.12.006
Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science*, *27*(11), 1451–1460.
Spaulding, S. (2020). What is mindreading? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(3), e1523.
Spaulding, S. (2023). Phenomenology of social explanation. *Phenomenology and the Cognitive Sciences*, *22*(3), 637–653.
Sulik, J., van Paridon, J., & Lupyan, G. (2023). Explanations in the wild. *Cognition*, *237*, 105464.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168.
Tschaepe, M. D. (2009). Pragmatics and pragmatic considerations in explanation. *Contemporary Pragmatism*, *6*(2), 25.
Van Fraassen, B. (1988). The pragmatic theory of explanation. *Theories of Explanation*, *8*, 135–155.
Vasilyeva, N., Wilkenfeld, D., & Lombrozo, T. (2017). Contextual utility affects the perceived quality of explanations. *Psychonomic Bulletin & Review*, *24*, 1436–1450.
Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*(3), 470–477.

Appendix

After collecting data and before conducting any inferential analyses, we performed a simulation-based power analysis approach (Brysbaert & Stevens, 2018; Kumle et al., 2021) to ensure sufficient statistical power (≥ 0.8) to detect the main effect of the condition.

Experiments 1a, 1b, 2a, and 2b

For each experiment, we simulated 1000 datasets matching our experimental design. Data for each experimental condition was simulated by extracting the probability distribution of ratings from our actual data. Using the *sample*() function, we generated synthetic data for

each condition (matching the number of the observations in the actual datasets), with the sampling probabilities specified by these extracted distributions. The simulated values and their corresponding condition labels were combined into a data frame, and this process was repeated for each condition to build the full simulated datasets. The *t*-value for the fixed effect was extracted to determine whether they exceeded the critical value ($|t| \geq 2$). The proportion of simulations where the critical effects were significant provided an estimate of the statistical power of the main effect in our model. The simulations evidenced that both Experiments 1a and 1b had enough power to test our hypothesis; the power of the critical effect was estimated at 80% for Experiment 1a, and at 100% in Experiment 1b (alpha level of 0.05). In Experiments 2a and 2b, the power reached 100%.

Experiment 3

We simulated 1000 datasets matching our experimental design. For each condition and category, we simulated binary outcome data (using the *rbinom*() function) reflecting the presence or absence of each category in participants' responses, based on the observed probabilities and sample sizes. We then fitted a generalized linear mixed-effects binomial regression to each simulated dataset, including fixed effects for condition (satisfied vs. not satisfied), category, and their interaction, as well as random intercepts and slopes (condition and category) for participants and items. In this model, as well as in the model for the actual data, we set the type of response "Category" as the intercept, since this was the type of response of interest in Experiment 3. Thus, the main effect of condition responds to the difference in the presence of "Category" as response type when the characters were satisfied with the explanation versus when they were not satisfied. Furthermore, the interaction between condition and response type reflects the difference in the presence of "Category" and each of the other answer type under the two experimental conditions—in other words, each interaction effect reflects a 2x2 contrast with Category against other type, and the two levels of the condition independent variable (i.e., satisfied vs. not satisfied). Based on 1000 simulations with a sample size of 33 participants, the power analysis of the critical effects revealed that, except for one (the interaction between condition and the difference between Category and Function, power = .6), all critical contrasts exhibited an estimated power of 95%.

Experiment 4

We simulated 1000 datasets matching our experimental design. For each simulation, participant responses were generated using random normal distributions (*rnorm()* function) corresponding to each combination of condition and CR score. The simulated ratings were rounded to the nearest integer and constrained within the 1–7 scale to reflect the actual response format. In each simulated dataset, we assigned participants to conditions and CRT levels based on the observed sample sizes. We then fitted the same linear mixed-effects model as specified above to each simulated dataset. Analysis of the simulated dataset showed that the power for the critical contrast (i.e., Interaction between CR and Tautology-Choice differences), with a sample size of 132 participants, exhibited an estimated power of 99.5%.