Cognition 240 (2023) 105586

Contents lists available at ScienceDirect

Cognition

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

Original articles

The impact of explanations as communicative acts on belief in a claim: The role of source reliability☆

Marko Tešić *, Ulrike Hahn

Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

ARTICLE INFO

Keywords:
Explanation
Reliability
Confidence
Belief revision
Trust
Pragmatics

ABSTRACT

Providing an explanation is a communicative act. It involves an explainee, a person who receives an explanation, and an explainer, a person (or sometimes a machine) who provides an explanation. The majority of research on explanation has focused on how explanations alter explainees' beliefs. However, one general feature of communicative acts is that they also provide information about the speaker (explainer). Work on argumentation suggests that the speaker's reliability interacts with the content of the speaker's message and has a significant impact on argument strength. In five experiments we explore the interplay between explanation, the explainee's confidence in what is being explained, and the explainer's reliability. Experiment 1 replicates results from previous literature on the impact of explanations on an explainee's confidence in what is being explained using real-world explanations. Experiments 2 and 3 show that providing an explanation not only impacts the explainee's confidence about what is being explained but also influences beliefs about the reliability of the explainer. Additionally, the two experiments demonstrate that the impact of explanation on the explainee's confidence is mediated by the reliability of the explainer. In Experiment 4, we experimentally manipulated the explainer's reliability and found that both the explainer's reliability and whether or not an explanation was provided have a significant effect on the explainee's confidence in what is being explained. In Experiment 5, we observed an interaction between providing an explanation and the explainer's reliability. Specifically, we found that providing an explanation has a significantly greater impact on the explainee's confidence in what is being explained when the explainer's reliability is low compared to when that reliability is high. Throughout the study we point to the important impact of background knowledge, warranting further studies on this matter.

1. Introduction

Explanations are ubiquitous. They are sought by both children and adults, in everyday as well as professional contexts, and they play an integral part in science. We often find ourselves and others asking why certain events have occurred: why is the road closed?; why is the baby crying?; why is the dog barking?; why is the Zoom meeting not starting?; why is Amazon recommending I buy this book? A doctor may ask why a child is in pain. A lawyer may question why a suspect is accused of a crime. A scientist may ask why is a certain theory a good explanation of some phenomena. The propositions that address these requests are explanations (Lombrozo, 2012). For instance, an explanation for why a child is in pain could be that they pulled a muscle, and an explanation for why a Zoom meeting has not started could be that the host forgot about it.

Research has consistently shown that providing an explanation for an event or a phenomenon increases confidence in it (Koehler, 1991; Lombrozo, 2006; Pennington & Hastie, 1993). Research has also pointed to specific features of explanations that make them more likely to be believed, such as simplicity and breadth (Lagnado, 1994; Lombrozo, 2006, 2007; Read & Marcus-Newhall, 1993; Thagard, 1989). However, the role of the source of explanations (i.e., the explainer) has been relatively underexplored in this area.

Explanations are communicative acts: it is often the case that someone (the explainer) provides an explanation to someone else (the explainee) (Hilton, 1990). For instance, a doctor (the explainer) may

☆ This research has been supported by the Alexander von Humboldt Foundation and the Royal Academy of Engineering UK IC fellowship.
* Corresponding author.
E-mail addresses: marko.tesic375@gmail.com (M. Tešić), u.hahn@bbk.ac.uk (U. Hahn).

https://doi.org/10.1016/j.cognition.2023.105586
Received 11 April 2022; Received in revised form 25 July 2023; Accepted 31 July 2023

offer an explanation of their diagnosis to a patient (the explainee), or a recommender system (the explainer) can provide an explanation for why it recommends a certain book to the user (the explainee). It is therefore reasonable to assume that when evaluating the impact of an explanation, people go beyond just examining its content and also take into account the source of the explanation, particularly their reliability.

Source reliability has been a prominent research topic in related areas such as argumentation, causal decision-making and reasoning (Collins, Hahn, von Gerber, & Olsson, 2018; Hahn, Harris, & Corner, 2009, 2016; Harris, Hahn, Madsen, & Hsu, 2016; Jarvstad & Hahn, 2011; Madsen, Hahn, & Pilditch, 2018, 2020; Merdes, Von Sydow, & Hahn, 2020). This body of research has clearly highlighted the interplay between the content of an argument that the speaker is putting forward and the speaker's reliability and the importance of considering the reliability of the source. Given the close relationship between explanation and arguments,[1] one would then expect that the reliability of an explainer plays an important role in how explanations impact our confidence in what is being explained.

In this paper we explore the role of an explainer's reliability in assessing the impact of explanations on our confidence in what is being explained. We start the following sections by discussing the definition and role of explanations (Section 1.1) and reviewing the literature on the effects of providing explanations (Section 1.2). We then further motivate the current research by pointing to the communicative aspects of explanations such as the reliability of an explainer (Section 1.3) and specify the kinds of explanations we focus on in this paper (Section 1.4). We close the introduction with an overview of the experiments presented in this paper (Section 1.5).

1.1. Definition and role of explanations

Defining explanation has proven to be a challenging task (Lombrozo, 2012). Explanations have been understood as answers to how- and why-questions, as well as judgments about why an outcome occurred or as hypotheses that include causes of what is being explained. These diverse ways of delineating explanation point to different aspects of an explanation. Generally speaking, however, explanations can be understood as propositions that address a request for an explanation (Lombrozo, 2012). For instance, in the case of the inference scheme called 'the inference of the best explanation' (IBE) the aim is to find the best explanation for evidence/facts (i.e. things we fully believe are true and/or are widely accepted to be true). Once such an explanation is found it then warrants the conclusion suggested by the explanation. For example, imagine one evening you leave a piece of cheese in your basement. The next morning, you find that except for a few crumbs the cheese is gone, and you notice a small hole in the bottom of the wall (facts). The best explanation for these observations is that a mouse was in your basement during the night (the example is due to Van Fraassen, 1980, pp. 19–20). The fact that this is the best explanation is then taken to license the belief that the explanation is, in fact, true.

Often, however, we also seek to explain claims or hypotheses about the world for which we are not entirely certain, or that are not widely accepted as true and whose truth (or falsehood) we are aiming to establish. For example, in court proceedings the prosecution may claim that the defendant killed person *A*. The prosecution would then present a case or make an argument that further elaborates on why they believe the defendant killed *A*, typically by elaborating on the connection between the evidence presented to the court and the claim that the defendant killed *A*. This elaboration serves as an explanation for that claim. In a scientific context, a scientist may wish to explain the hypothesis that cows produce more milk when they listen to classical music (Holden, 2001). An explanation could involve providing a causal mechanism that links listening to classical music and cows' milk production such as: classical music relaxes cows and relaxed cows produce more milk (Glassner, Weinstock, & Neuman, 2005). As illustrated by these two examples, explanations of claims/hypotheses often involve providing a causal mechanism or a reasoning process regarding how two (or more) factors are related (e.g. how the defendants actions and person *A*'s death are related or how classical music and cows' milk production are connected).

Explanations of claims have been extensively explored by numerous researchers. For instance, Toulmin in his book *The uses of argument* introduces his highly influential argument framework (Toulmin, 1958/2003). There, he differentiates between claims, data, and warrants. Warrants would be explanations that make obvious the connection between what is being argued for in the claim and certain observations. Thagard (1989) views explanations as the key to providing coherence between hypotheses (claims), while Antaki and Leudar (1992) considers explanations as providing support for claims. Additionally, Brem and Rips (2000) explored whether people can distinguish between explanation and evidence. They distinguish between the claim ('a proposition whose truth value we are attempting to establish'), evidence (data), and explanation that can provide support for the claim by providing a (causal) bridge between the data and the claim. For instance, one could argue that welfare recipients have difficulty getting off public aid (data) because they lack job skills (claim). However, it is not necessarily obvious how the lack of skills could lead to the difficulty of getting off public aid. To that end one could provide an explanation such as 'Job skills increase a person's chances of landing a well-paid job, which in turn supplies them with enough money to give up welfare checks'. This explanation further elucidates the relationship and inference process between the claim and data, thereby serving as a bridge between them. Similarly, Glassner et al. (2005) view explanations as providing a theoretical basis (e.g. causal mechanism) for the phenomena described in the claim.

In this paper, we investigate the effects of explanations of claims/hypotheses on our confidence in the truth of these claims/hypotheses. We take claims or hypotheses to suggest a (causal) relationship between phenomena (e.g. between person *A*'s death and the defendant's actions or between cows' milk production and classical music). Explanations serve as a means to further clarify that relationship by, for example, providing a (causal) mechanism.

1.2. Effects of explanations

We have discussed what explanations are and how we can conceptualize them, but what are some of the effects of explanations? Often, explanations are thought to increase the sense of understanding (Hahn, 2011; Hempel, 1965; Lombrozo, 2012): providing an explanation of a phenomenon increases the recipient's sense of understanding of what has been explained. However, it has been pointed out by Trout (2002, 2008) that this increase in the sense of understanding may not necessarily translate into an increase in actual understanding.

Furthermore, research has found that generating explanations (even incorrect ones), rather than just receiving explanations, has a beneficial impact on learning (Lombrozo, 2012). Explanations can also influence the perceptions of normality: finding plausible explanations of, for instance, patients' behavior can lead to perceiving the patients as more 'normal' than when such an explanation was lacking (Ahn, Novick, & Kim, 2003). Additionally, providing an explanation of a hypothetical outcome or of a past event that we are uncertain about whether it happened increases the subjective likelihood of the hypothetical outcome occurring in the future and of the event that might have occurred in the past (Koehler, 1991, 1994; Ross, Lepper, Strack, & Steinmetz,

[1] For example, Hempel (1965) argued that explanations *are* arguments. Furthermore, one of the most influential reasoning schemes is what is often referred to as 'inference to the best explanation' (Harman, 1965; Lipton, 2003). Despite the similarities between explanations and arguments, authors have also acknowledged important differences (Antaki & Leudar, 1992; Hahn, 2011; Walton, 2004a).

1977; Sherman, Zehner, Johnson, & Hirt, 1983). Relatedly, Davoodi and Lombrozo (2022) find that asserting ignorance of an explanation may impact the confidence in the truth of a claim which differs across different domains.

These are only some of the effects of explanations. However, the effect of explanations that we will be focusing on in this paper is their impact on the recipients' confidence in the claims/hypotheses. We will review the literature regarding this effect next.

1.2.1. Effects of explanations on confidence

Inference to the best explanation (IBE) nicely illustrates how explanations impact confidence. The fact that a proposition is the best explanation (i.e. it has the highest explanatory goodness compared to other rival explanations) increases the subjective probability (or confidence) assigned to that hypothesis (Douven, 2013, 2017; Harman, 1965; Lipton, 2003; Psillos, 2000). Douven and Schupbach (2015) suggest this is empirically also the case and that people judge a hypothesis to be more likely (i.e. they are more confident in the hypothesis being true) if the explanatory goodness of that hypothesis also increases. Some of the factors that influence the explanatory goodness of a hypothesis and thus increase its subjective probability are simplicity (Lagnado, 1994; Lombrozo, 2007; Read & Marcus-Newhall, 1993; Thagard, 1978), breadth (Lombrozo, 2016; Read & Marcus-Newhall, 1993; Thagard, 1989), consistency with prior knowledge (Thagard, 1989) and coherence (Pennington & Hastie, 1993; Thagard, 1989).

In the context of explaining hypotheses/claims it has been found that asking people to provide an explanation as to whether a particular property is true or false changes their perceived/subjective likelihood of that property (Lombrozo, 2006). For instance, when participants were asked to explain the relationship between two variables A and B (e.g. why risky people (A) are better firefighter (B)), their subjective estimates of the relationship significantly increased compared to both the control group, which was not prompted to explain the relationship (Anderson, Lepper, & Ross, 1980) and their previous estimates when they were not asked to explain the relationship (Anderson & Sechler, 1986). Furthermore, people's confidence that the event for which they provided an explanation would occur in the future was higher compared to when no explanation was provided (Koehler, 1991).

Similarly, Thagard (1989) argues that a hypothesis that the accused murdered the victim will be more plausible if we find reasons for why the accused was motivated to kill the victim (explanation). Empirical work by Pennington and Hastie (1993) supports this idea. They find that the story summary (explanation), which is the interpretation of the evidence (data) presented in a narrative story form, impacts the confidence in a juror's decision: the better the story (explanation) the greater the impact on confidence. Finally, Brem and Rips (2000) also argue that the perceived probability of a claim may be increased as a result of there being an explanation for that claim.

In summary, both theoretical and experimental works suggest that ether providing or receiving an explanation will result in increased confidence in the claim across different contexts and notions of explanation.

1.3. Explanations as communicative acts

Explanations have an important social dimension. They often occur between individuals who are trying to communicate understanding (Keil, 2006), and they typically take the form of a conversation where "*[s]omeone* explains *something* to *someone*" (Hilton, 1990, p. 65, original emphasis). Explanations are then in their essence communicative acts that involve interpersonal exchange and include two parties: an explainer and an explainee. In line with the literature mentioned in Section 1.2 we would expect the explainee's confidence in a claim to be affected by the explainer's explanation. Specifically, we would expect that the act of the explainer providing an explanation would increase the explainee's confidence in the claim being explained.

The communicative dimension of explanations introduces additional factors that can influence confidence. We often rely on others (e.g. experts) to provide us with explanations about various phenomena. For example, experts may be called upon to explain to the general public why a particular virus is dangerous to the population. The fact that experts are providing us with an explanation may impact our confidence in the claim that the virus is dangerous. Moreover, explanations as communicative acts convey information about the speaker (the explainer) that could affect the explainee's confidence. One of the aspects of the explainer that could influence the explainee's confidence is the explainer's *reliability*.

The effects of the reliability of the source of information have been extensively explored, both theoretically and empirically, in the context of argumentation. Formal models of source reliability that aim to distill the impact of reliability on confidence that goes beyond the argument content have been proposed by Bovens and Hartmann (2003) and Olsson and Vallinder (2013) (for a detailed review see Merdes et al., 2020). Some of these models have been empirically tested. For instance, Hahn et al. (2009) varied both argument strength and the reliability of the sources and found that both argument strength and the reliability of the source affected the participants' confidence in the arguments, with an interaction between the two, which was in line with some of the formal models. Similarly, Harris et al. (2016) found that greater expertise and reliability increase the impact on one's confidence in claims (see also Collins & Hahn, 2019; Collins et al., 2018; Hahn et al., 2016; Hahn, Oaksford, & Harris, 2013; Walton, 2007). Furthermore, Jarvstad and Hahn (2011) observed that perceived reliability can be influenced by evidence (data) or the report from the source, with a more likely statement being judged to come from a more reliable source.

In this paper we aim to explore the impact of the explainer's reliability on confidence in the claim both when there is an explanation for the claim and when such an explanation is missing. Consistent with the argumentation literature on argument content and reliability, we expect to find differing impacts of reliability when the explanation for the claim is provided compared to when such an explanation is not provided.

To the best of our knowledge, the explanation literature has not experimentally manipulated the impact of reliability on confidence in claims to explore its effects. However, some exploratory analyses have been conducted. For instance, Zemla, Sloman, Bechlivanidis, and Lagnado (2017) explore criteria that predict explanation quality and found that expertise is one of the factors significantly predicting explanation quality, with higher (perceived) expertise being related to better quality explanations. This potentially suggests that expertise positively influences confidence: the more the explainer is perceived as an expert the higher the confidence in the claim. Wilkenfeld, Plunkett, and Lombrozo (2016) identified a correlation between explanatory depth (the level of detail in explanations regarding a specific fact) and people's willingness to defer to the source of that explanation: the higher the explanatory depth of a person's explanation, the more willing individuals are to defer to the source of the explanation for domain-related questions. Relatedly, Keil, Stein, Webb, Billings, and Rozenblit (2008) found that from a young age people's understanding of domains of knowledge plays a important role in determining whether an expert knowledgeable about X would also know and provide explanations about Y (a fact from the same domain as X) or about Z (a different domain compared to facts X and Y). These studies suggest a relationship between explainer's reliability and people's perception of (the goodness of) explanations. We aim to experimentally explore the impact of the explainer's reliability on the confidence in the claims as well as how providing an explanation and differing levels of reliability combine to influence the confidence in the claims. By systematically examining these factors, we can gain a better understanding of the complex interplay between explanation, explainer's reliability, and the resulting impact on confidence.

1.4. Everyday explanations

Before we proceed to describe the experimental exploration, a brief note on explanations used for this exploration. Some psychological investigations on explanations mentioned earlier have been derived from the philosophical literature (e.g. from Hempel & Oppenheim, 1948; Kitcher, 1989; Lipton, 2003) and have studied how lay people perceive the aspects of explanations that the philosophical literature has considered important in judging scientific explanations (e.g. Lombrozo, 2007). These empirical studies have, for good reasons, often employed short and simple explanations with a minimal causal structure, sometimes, with a single cause and effect (for an overview see Lombrozo, 2012).

More recently psychologists have looked into everyday explanations to explore the sets of criteria used to judge the explanatory goodness of these kinds of explanations (e.g. Bechlivanidis, Lagnado, Zemla, & Sloman, 2017; Zemla et al., 2017). The aim of these studies was to investigate whether the criteria for evaluating scientific explanations also apply to everyday explanations. Furthermore, exploring the explanatory criteria in the context of real-world explanations would arguably provide a more ecologically valid understanding of these criteria. Everyday explanations are more nuanced than the experimental stimuli often used in psychological studies in that they include multiple causes that can form chains of causal reasoning. Additionally, everyday explanations are naturally embedded in conversational forms, allowing for a more natural exploration of the communicative aspects of explanations and the impact of reliability on confidence. In this paper, we thus employ everyday explanations as materials for our empirical investigation of the impact of explanations and reliability.

1.5. Overview of experiments

The aim of this paper is to explore the relationship between (everyday) explanations of claims, the reliability of an explainer, and the explainee's confidence in a claim. Experiment 1 tested the impact of explanations on confidence in a claim without considering of the social aspects of explanations, such as the reliability of the explainer. The goal of this experiment was to replicate the findings from previous literature regarding the effects of explanations on confidence with new experimential materials. Experiments 2 and 3 included the social aspects of explanation and tested the impact of explanations not only on the confidence in the claim, but also on the reliability of the explainer. The aim here was to explore whether providing an explanation affects the perceived reliability of the explainer and whether the effect of explanations on confidence is mediated by the explainer's reliability. Experiment 4 explored the impact of both the explanation and the explainer's reliability on the explainee's confidence in a claim, aiming to investigate the potential causal impact of the explainer's reliability on confidence in the claim. Experiment 5 explored the interaction between the explainer's reliability and explanation with the goal of further elucidating the mediating effects of the explainer's reliability on the explainee's confidence in the claim.

2. Experiment 1

The aim of this experiment was to replicate the findings of previous studies on the effects of explanation on people's confidence using real-world explanations as stimuli. Following these studies, we expected that adding an explanation would increase people's confidence in the claim.

2.1. Methods

2.1.1. Participants & design

A total of 130 participants ($N_{\text{FEMALE}} = 87$, $M_{\text{AGE}} = 33.8$ years) were recruited from Prolific Academic (www.prolific.co). All participants were native English speakers residing in the UK, the US, or Canada with approval ratings of 95% or higher. They all gave informed consent and were paid £5 an hour rate for participating in the present study, which took on average 10.5 min to complete.

Participants were randomly assigned to either the control group where no explanation of the claim was provided ($N = 66$) or the treatment group where an explanation was provided ($N = 64$).

2.1.2. Materials

In all experiments presented in this paper we employed the following scenarios adapted from Zemla et al. (2017), who used Reddit's *Explain Like I'm Five* (Eli5; www.reddit.com/r/explainlikeimfive), Wikipedia, and HowThingsWork.com to source the stimuli and explanations. These platforms are widely accessible to the general public and the issues addressed on these platforms are often aimed at the general population, covering a wide range of phenomena that one can encounter in daily life. The scenarios were picked from three different domains: public health, social policy, and history. The scenarios were chosen with the idea that the general public would be interested in them. All scenarios had the same format. The first paragraph started with an introduction of up to two sentences setting up the context and a question seeking an explanation. The second paragraph described a claim/hypothesis (no explanation condition) or it described a claim/hypothesis and an explanation that provided a clarification for what is presented in the claim (the explanation condition). Lastly, participants were asked a question that elicited their confidence estimates in the claim. This format was very similar to the one adopted by Brem and Rips (2000, Experiment 2).

For example, the *Black Death* scenario looked as follows. Note that the text of the scenarios and the questions were the same for both the no explanation and the explanation conditions, except for the part in the curly brackets ({}) that appeared only in the explanation condition. The text in the square brackets did not appear in either condition and is added here to point to the functions of the different parts of the scenario.

> Millions of people died from the Black Death in the 14th century. **How did the Black Death come to an end?** [*introduction*]
>
> One popular belief is that the Black Death subsided mostly through the use of quarantines. [*claim*] {According to this belief, people mostly stayed out of the path of infected individuals, rats, and fleas. The uninfected would typically remain in their homes and only leave when it was necessary. Those with the financial resources would traditionally escape to the country, far away from the Black Death-infested cities. [*explanation*]}
>
> **Q**. How confident are you that the Black Death came to an end through the use of quarantines? [*a question eliciting participants' confidence in the claim*]

The other four scenarios concerned the increase in China's population despite the one-child policy, the way medical practitioners contract Ebola, Switzerland's armed neutrality during World War II, and the way vaccines build immunity.

Zemla et al. (2017) experimentally explored the quality of explanations used in this study by asking participants to rank the explanations on a 7-point Likert scale from '1–Strongly disagree' to '7–Strongly agree' in response to the statement 'This is a good explanation'. For all five explanations used in the five scenarios they found that participants rated them well above average in quality: the explanation from the

Switzerland scenario had an average rating of 5.4, from the Ebola scenario 6.4, from the China scenario 5.1, from the Vaccination scenario 5.7, and from the Black Death scenario 5.9.

For the full materials used in all five experiments presented in this paper visit the Open Science Framework, at https://osf.io/wkhdc/. The data collected in the five experiments are available via the same link.

2.1.3. Procedure

After giving informed consent and basic demographic information, participants were shown the following instructions:

> **WELCOME!**
>
> You will now be presented with 5 explanations of 5 events and phenomena found in the real world and required to answer some questions related to the explanations.
>
> Please make sure you read all the information carefully before answering the questions.

Following these instructions, participants were presented with the five scenarios and questions related to these scenarios. The order in which the scenarios were presented was randomized for each participant. Each scenario was presented on two pages. On the first page was the main text of the scenario. On the second page, the text of the scenario was repeated as a reminder and participants were asked two questions: one about their confidence in the claim and another to explain their reasoning regarding how they arrived at their confidence estimate. The second question was asked to gain additional insight into participants' reasoning.

For example, the *Black Death* scenario had the following text on the first page (the additional text that appeared only in the explanation condition is in curly brackets):

> Millions of people died from the Black Death in the 14th century. **How did the Black Death come to an end?**
>
> One popular belief is that the Black Death subsided mostly through the use of quarantines. {According to this belief, people mostly stayed out of the path of infected individuals, rats, and fleas. The uninfected would typically remain in their homes and only leave when it was necessary. Those with the financial resources would traditionally escape to the country, far away from the Black Death-infested cities.}

On the second page, the scenario was repeated and the questions related to the scenario were asked:

> **Reminder:**
>
> Millions of people died from the Black Death in the 14th century. **How did the Black Death come to an end?**
>
> One popular belief is that the Black Death subsided mostly through the use of quarantines. {According to this belief, people mostly stayed out of the path of infected individuals, rats, and fleas. The uninfected would typically remain in their homes and only leave when it was necessary. Those with the financial resources would traditionally escape to the country, far away from the Black Death-infested cities.}
>
> ***Please answer the following question.***
>
> **Q.** How confident are you that the Black Death came to an end through the use of quarantines?
>
> [A slider eliciting confidence (%) on a scale from 0% to 100%.]
>
> **R.** Please explain your reasoning for your answer to the question in the box below.
>
> [A text box.]

A percentage scale from 0% to 100% was used to elicit participants' confidence estimates in the claims in question. A free format text box was used to ask participants to explain their reasoning for the estimates they provided. Lastly, after the participants answered questions for all five scenarios they received debriefing information.

2.2. Results & discussion

To analyze the data,[2] we built a linear mixed effect model (LMM) using the "lme4" package in R (Bates, Mächler, Bolker, & Walker, 2014). The only fixed effect was group (with two levels: no explanation and explanation). The only random effect was the intercept for participants. There was no random slope from the participant as the design was fully between. No random intercept for scenarios was used as the number of scenarios was low (i.e. 5) and including the scenarios as a random intercept could have led to reduced power of the experiment (see Judd, Westfall, & Kenny, 2017; Singmann & Kellen, 2019). Further, a random slope for scenarios was not included as it led to a singular fit model, implying that the variance of this random effect was (close to) zero.

The LMM indicated that confidence estimates in the explanation group (Estimated Marginal Mean (EMM) = 70.02) were significantly higher than in the no explanation group ($EMM = 58.91$); $t(128) = 3.96$, $p < .001$ (see Fig. 1(a)). Furthermore, the inclusion of the predictor for the group in the model led to a significant improvement in model fit ($\chi^2(1) = 15$, $p < .001$), compared to just having an intercept as a predictor. This result is in line with the previous literature and provides further support that people's confidence in claims is higher when an explanation is provided.

The effect of explanation was not only observed overall, but also within each of the five scenarios. Fig. 1(b) shows that participants' confidence estimates in the explanation conditions were higher in all scenarios, suggesting that the effect was not driven by specific scenarios. Further, the mean confidence estimates were similar across the scenarios (in the respective explanation/no explanation conditions), except in the *Vaccination* scenario (particularly in the explanation condition of that scenario). From participants' textual answers where they provided reasons for choosing a specific confidence estimate in this scenario we noticed that a number of participants have said that the claim and explanation agreed with what they already knew about vaccination, which led them to provide higher estimates in both the no explanation and explanation conditions of this scenario. This finding hints at the importance of background knowledge in judging people's confidence in claims supported by explanations. In the next experiments we find further support for the effects of background knowledge.

3. Experiment 2

In Experiment 1 participants were asked to provide their confidence estimates in claims without information about the source of these claims. However, since explanations are communicative acts they often involve a speaker (an explainer) and could potentially provide information about the speaker's reliability. Explanations then not only affect confidence in the claims but could also influence the perceived reliability of the speaker providing the explanation.

The aim of this experiment was to explore the impact of an explanation on both the confidence in a claim and the perceived reliability of the source that provided the claim and explanation.

3.1. Methods

3.1.1. Participants & design

A total of 52 participants ($N_{\text{FEMALE}} = 31$, $M_{\text{AGE}} = 32.6$) were recruited from Prolific Academic (www.prolific.co). The selection criteria and remuneration rate per hour were the same as in Experiment 1. Participants took on average 16.4 min to complete the experiment.

Participants were randomly assigned to either the control group where no explanation for the claim was provided ($N = 24$) or the treatment group where an explanation was provided ($N = 28$). All participants were asked to provide estimates on two dependent variables: confidence and reliability.

[2] In light of the COVID-19 pandemic, it is worth pointing out that the data for Experiments 1, 2, 3, and 4 were collected in the period between June and September 2019. The data for Experiment 5 was collected in January 2023.

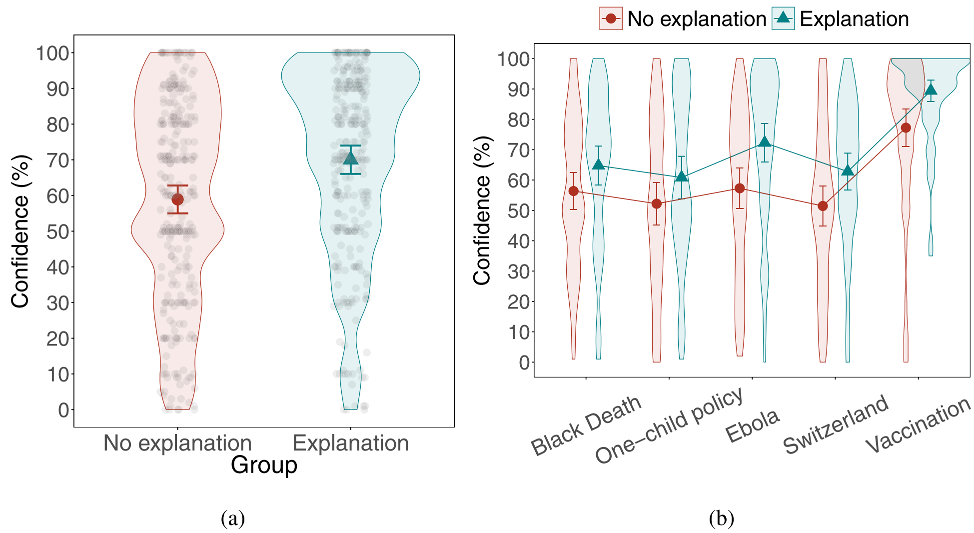


Fig. 1. (a) The estimated marginal means (EMMs) from the LMM built for Experiment 1 with 95% confidence intervals. Gray points are raw data values (jittered along the x-axis for visibility) with violin plots showing the frequency of the raw data. (b) The observed data means (with 95% confidence intervals) and violin plots for each scenario broken down for each explanation conditions.

3.1.2. Materials

To explore the communicative aspect of explanations and their impact on the explainer's reliability we followed Hahn et al. (2009) who used a dialogue form to study the impact of reliability and the content of an argument on confidence in what is argued for. This format is also in line with Walton (2004b) who argues that a dialogue form is appropriate for explanations.

The same five scenarios from Experiment 1 were employed in this experiment and further adapted to fit the form of a dialogue between two people, an explainer and an explainee, where the explainer provided the claims and explanations. Such a format enables us to elicit not only participants' confidence estimates in claims but also their reliability estimates in the explainer as a source of the claims and explanations.

The adaptation of scenarios into dialogues was done in a similar manner as in Hahn et al. (2009). For example, the *Black Death* scenario was adapted in a way that it included two people, an explainer (Dave) and an explainee (Jimmy), where Jimmy is asking questions and Dave is trying to provide answers (the part in curly brackets appeared only in the explanation condition):

> Dave and Jimmy are part of a research group investigating devastating pandemics in human history. During a planning meeting they touched upon the Black Death.
>
> **Dave**: Millions of people died from the Black Death in the 14th century. I think our research project should in part focus on how the Black Death ended. It may give us some insight into how to deal with future pandemics.
>
> **Jimmy**: Yes, I agree. Do you already have an idea regarding how the Black Death came to an end?
>
> **Dave**: I think the Black Death subsided mostly through the use of quarantines.
>
> {**Jimmy**: How so?
>
> **Dave**: People mostly stayed out of the path of infected individuals, rats, and fleas. The uninfected would typically remain in their homes and only leave when it was necessary. Those with the financial resources would traditionally escape to the country, far away from the Black Death-infested cities.}

The other scenarios were adapted in a similar way.

3.1.3. Procedure

The procedure for this experiment was similar to the procedure for Experiment 1 in that the welcome page was shown after the participants gave informed consent and demographic information, and each scenario was presented in a random order on two pages. The difference lies in that participants now answered two questions in each scenario: one about the confidence in the claim and one about the reliability of the explainer. For example, after being shown and reminded of the *Black Death* scenario the participants were asked:

> **Q1**. How confident are you that the Black Death came to an end through the use of quarantines?
>
> [A slider eliciting confidence (%) on a scale from 0% to 100%.]
>
> **Q2**. How reliable do you think **Dave** is as a source of information regarding the end of the Black Death?
>
> [A slider eliciting reliability (%) on a scale from 0% to 100%.]

For both the confidence questions and the reliability questions participants were asked to move the slider which was on the scale from 0% to 100%. Both the confidence and the reliability questions were followed by free format type text boxes where participants could explain their reasoning for selecting certain confidence/reliability estimates. Finally, participants received debriefing information.

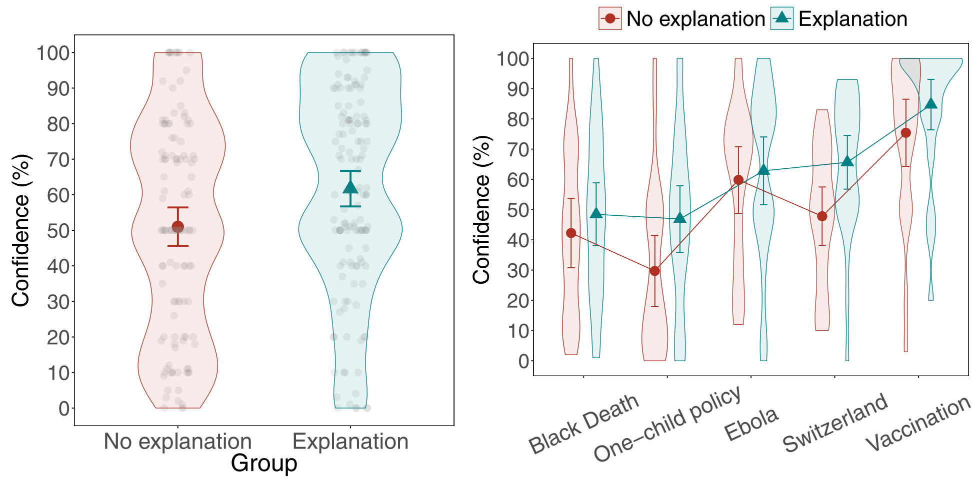
3.2. Results & discussion

Separate analyses were conducted for each dependent variable.

3.2.1. Confidence

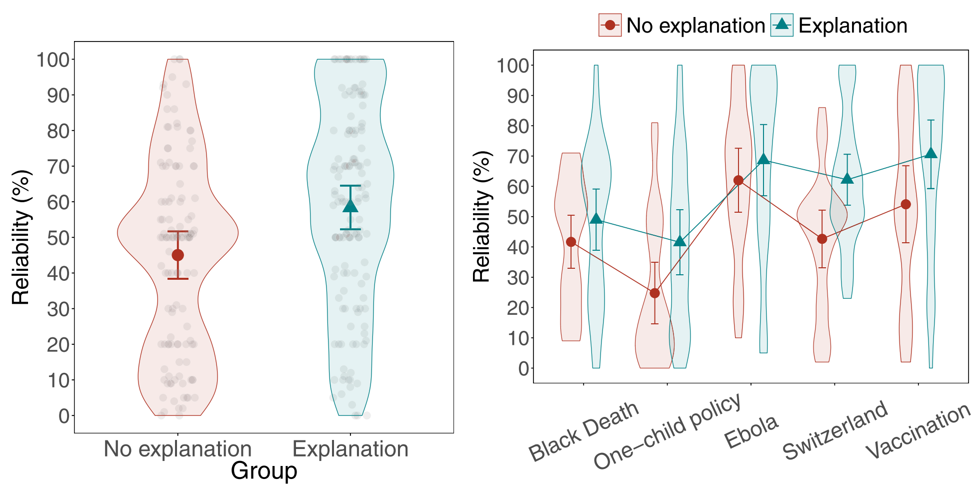
The LMM with the same random effects structure as in Experiment 1 indicated that confidence estimates in the explanation group ($EMM = 61.7$) were significantly higher than in the no explanation group ($EMM = 51$); $t(50) = 2.91$, $p = .005$ (see Fig. 2(a)). Further, the inclusion of the predictor for the group in the model led to a significant improvement in model fit ($\chi^2(1) = 8.1$, $p = .004$), compared to just having an intercept as a predictor.

This result is in line with the finding from Experiment 1 and is consistent with previous literature. The similar trend was found in all five scenarios (see Fig. 2(b)).



(a) (b)

Fig. 2. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *confidence estimates* in Experiment 2. Gray points are raw data values (jittered along the x-axis for visibility) with violin plots showing the frequency of the raw data. (b) The observed data means (with 95% confidence intervals) and violin plots for each scenario broken down for each explanation conditions.



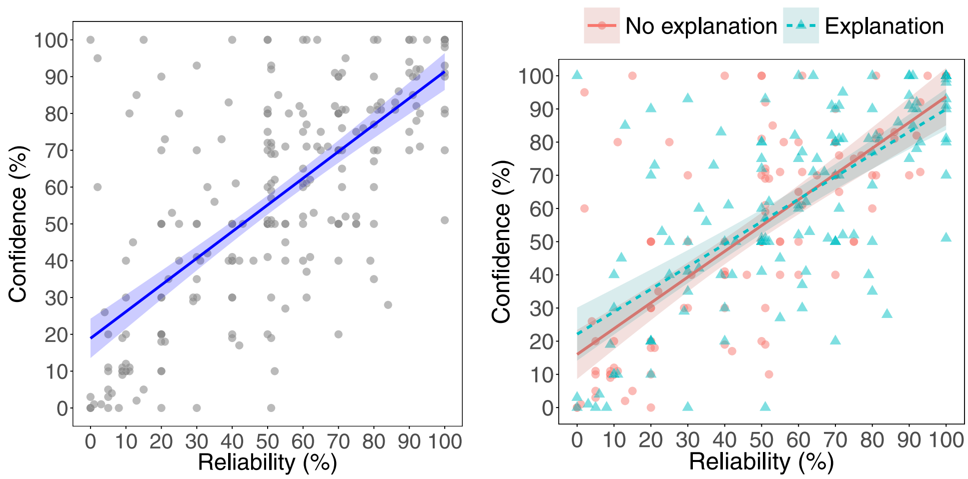
(a) (b)

Fig. 3. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *reliability estimates* in Experiment 2. Gray points are raw data values (jittered along the x-axis for visibility) with violin plots showing the frequency of the raw data. (b) The observed data means (with 95% confidence intervals) and violin plots for each scenario broken down for each explanation conditions.

3.2.2. Reliability

The LMM indicated that reliability estimates in the explanation group ($EMM = 58.39$) were significantly higher than in the no explanation group ($EMM = 45.03$); $t(50) = 2.97$, $p = .005$ (see Fig. 3(a)). Further, the inclusion of the predictor for the group in the model led to a significant improvement in model fit ($\chi^2(1) = 8.4$, $p = .004$), compared to just having an intercept as a predictor. This suggests that providing an explanation can also have an effect on our perceptions of the reliability of an explainer, not just on our confidence in what is explained (i.e. the claim).

We again found a similar general trend across the five scenarios with some variations in the magnitude (see Fig. 3(b)). These variations seem to correspond to the level of expertise the explainer has. For instance, in the *Ebola* and the *Black Death* scenarios the explainers were a medical practitioner and a member of a research group investigating devastating pandemics in human history respectively (see full Materials stored on the OSF page for this paper). Plausibly both of these explainers could be considered experts in their fields implying that their reliability is high in the context of these scenarios. In contrast, in *Vaccination* and *Switzerland* scenarios the explainers were students discussing a student project whose reliability in these contexts is arguably low. In the *One-child policy* scenario no information on the explainer's occupational or professional background was provided suggesting no specific level of expertise. However, the scenario context seems to suggest that the explainer and the explainee have only touched upon China's one-child policy in a (casual) conversation suggesting potentially that the explainer is a non-expert. These different levels of expertise (expert vs. non-expert) seem to correspond to the magnitude of the difference between the mean reliability estimates in each explanation condition: in the scenarios where the explainer is an expert (high reliability) it seems that the differences in mean reliability estimates between the two explanation conditions are smaller compared to these differences in the scenarios where a non-expert (low reliability) plays a role of an explainer. In Experiments 4 and 5 we experimentally manipulate the expertise of an explainer to further explore the impact of explanations when the explainer's reliability is at different levels.



(a) (b)

Fig. 4. (a) The raw data values of participants' reliability and confidence estimates from Experiment 2 and a linear regression model (with the 95% confidence band). (b) The same data and a linear regression model as in (a) broken down for each explanation condition.

3.2.3. Mediation analysis

A closer look at participants' estimates on the two dependent variables reveals a strong relationship between reliability and confidence in our data (Fig. 4(a)): Pearson's correlation $r = .7$, $t(258) = 16$, $p < .001$. One possible explanation of this relationship is that participants simply copied their confidence estimates into their reliability estimates (or vice versa) due to, potentially, their disengagement or misunderstanding of the task. This possibility is explored in Experiment 3.

Another possibility, however, is that reliability is mediating the effect of explanation on confidence that we found in both Experiment 1 and Experiment 2. The initial support for the mediation is readily found in Fig. 4(b) where we observe that the same strong relationship between reliability and confidence is preserved and unchanged when data is broken down for each explanation condition. We explore this possibility in more detail here.

Analyses in Sections 3.2.1 and 3.2.2 suggest that explanation has a significant effect on both confidence and reliability. Following Baron and Kenny (1986), to explore whether reliability mediates the effect of explanation on confidence we also built an LMM model with both explanation and reliability as predictors of confidence (and the same random effects structure as in the above models). If the effect of explanation on confidence in this model was reduced compared to when the only predictor of confidence was explanation (as in Section 3.2.1), then this would suggest that reliability is (partially or fully) mediating this effect. We found that when reliability is also included as one of the predictors of confidence, the effect of explanation on confidence disappears ($t(257) = 0.39$, $p = .69$) whilst the effect of reliability on confidence is highly significant ($t(257) = 15.4$, $p < .001$) (see Fig. 5 for a graphical summary of this meditation analysis). Using the "mediation" package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) we found that the mediation effect is significant ($p = .004$) and that reliability mediates around 90 percent of the association between explanation and confidence. This suggests that a large proportion of the effect that explanation has on confidence is mediated by reliability and that reliability may have a causal effect on confidence. This potential causal effect of reliability on confidence is further explored in Experiments 4 and 5 below.

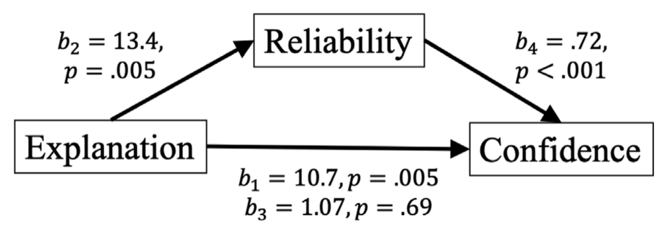


Fig. 5. Reliability as a mediator between explanation and confidence. b_1, with the related p-value, is the coefficient in the LMM with explanation as a predictor and confidence as a dependent variable (Section 3.2.1); b_2 is the coefficient in the LMM with explanation as a predictor and reliability as a dependent variable (Section 3.2.2); b_3 and b_4 are coefficients for explanation and reliability respectively in the LMM with explanation and reliability as predictors and confidence as a dependent variable (Section 3.2.3). In contrast to b_1, b_3 is minimal and non-significant which suggests that reliability (fully) mediates the effect of explanation on confidence.

4. Experiment 3

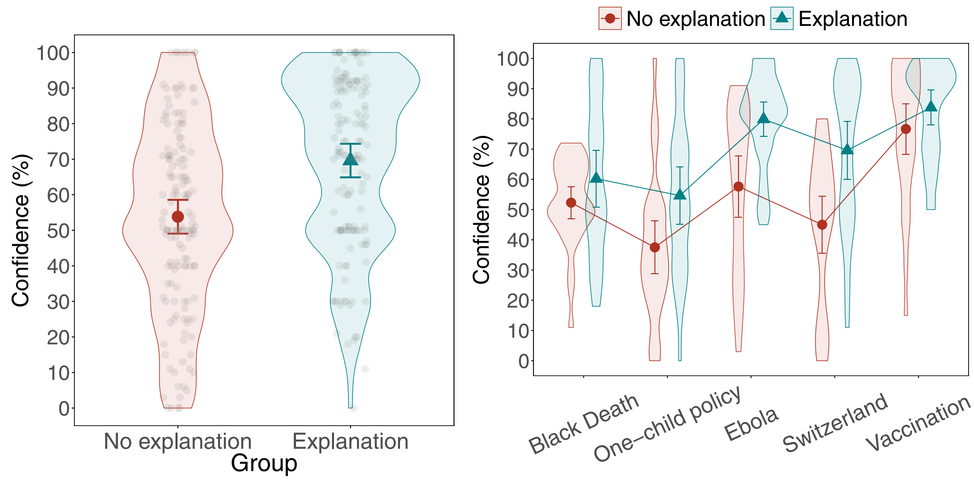
The aim of this experiment was to further explore the possibility suggested by the findings from Experiment 2: namely, that the strong relationship between participants' confidence and reliability estimates in Experiment 2 was due to participants simply copying their confidence estimates into their reliability estimates (or vice versa). To investigate this, instead of eliciting both confidence and reliability estimates from all participants, we asked them to provide either the confidence or the reliability estimates, but not both. If Experiment 3's results are similar to those in Experiment 2, it would provide more assurance that the effect of explanation on confidence and reliability is genuine, and that reliability plays a role in mediating the effect of explanation on confidence.

4.1. Methods

4.1.1. Participants & design

A total of 121 participants ($N_{FEMALE} = 81$, two participants identified as neither male nor female, $M_{AGE} = 34.7$ years) were recruited from Prolific Academic (www.prolific.co). The selection criteria and the remuneration rate per hour were the same as in the previous experiments. Participants took on average 10 min to complete the experiment.

The design of Experiment 3 was similar to the design of Experiment 2, with the difference that participants in both the no explanation and explanation conditions were asked to provide either their confidence rating in the claim or their reliability rating of the explainer, but not both. As a result, participants were randomly allocated to one of four groups: a no explanation group where only confidence rating was elicited ($N = 30$), a no explanation group where only reliability rating



(a) (b)

Fig. 6. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *confidence estimates* in Experiment 3. Gray points are raw data values (jittered along the x-axis for visibility) with violin plots showing the frequency of the raw data. (b) The observed data means (with 95% confidence intervals) and violin plots for each scenario broken down for each explanation conditions.

was elicited (N = 30), an explanation group where only confidence rating was elicited (N = 30), and an explanation group where only reliability rating was elicited (N = 31).

4.1.2. Materials

The same scenarios and questions as in Experiment 2 were used.

4.1.3. Procedure

The procedure was identical to Experiment 2, with the difference that participants were asked only one question rather than two: they were asked either the question about their confidence in the claim or the question about their perceived reliability in the explainer.

4.2. Results & discussion

4.2.1. Confidence

The LMM indicated that confidence estimates in the explanation group (EMM = 69.62) were significantly higher than in the no explanation group (EMM = 53.82); $t(58) = 4.73$, $p < .001$ (see Fig. 6(a)). Further, the inclusion of the predictor for the group in the model led to a significant improvement in model fit ($\chi^2(1) = 19.5$, $p < .001$), compared to just having an intercept as a predictor. This trend is also preserved in each of the scenarios (see Fig. 6(b)). These results are consistent with those from both Experiment 1 and Experiment 2 on confidence and they all follow the same trends.

4.2.2. Reliability

The LMM with reliability as a dependent variable showed that reliability estimates in the explanation group (EMM = 62.69) were significantly higher than in the no explanation group (EMM = 52.51); $t(59) = 2.33$, $p = .023$ (see Fig. 7(a)). Further, the inclusion of the predictor for the group in the model led to a significant improvement in model fit ($\chi^2(1) = 5.36$, $p = .021$), compared to just having an intercept as a predictor. These results also follow the same general trend as those in Experiment 2, suggesting that participants in Experiment 2 did not simply copy their confidence estimates into their reliability estimates and supporting the idea that the effect of explanation on reliability is genuine.

Zooming in on specific scenarios we found similar general trends, i.e. participants' reliability estimates were on average higher in the explanation condition than in the no explanation condition, except in the *Vaccination* scenario where the average reliability estimate in the explanation group was lower than in the no explanation group (see Fig. 7(b)). Looking into participants' textual explanations of their reasoning for the estimates they provided, we again found hints of the effects of background knowledge. Namely, 11 participants (out of 30) in the no explanation condition of the *Vaccination* scenario wrote that the claim agreed with their personal (background) knowledge of how vaccines work and all of them provided reliability estimates higher than 60% (this subgroup's average reliability estimate was 86%). Their typical explanations were 'His [explainer's] answer is what I would have said' or 'My understanding [of how vaccines work] is the same as his [explainer's].' The number of participants who provided explanations similar to these and pointed to their background knowledge was only 6 (out of 31) in the explanation condition and all their estimates were also higher than 60% (their average reliability estimate was 84.3%). This shows how (agreement with) people's background beliefs and knowledge can affect their reliability estimates of a person providing an explanation, sometimes even trumping the effects of explanation on reliability.

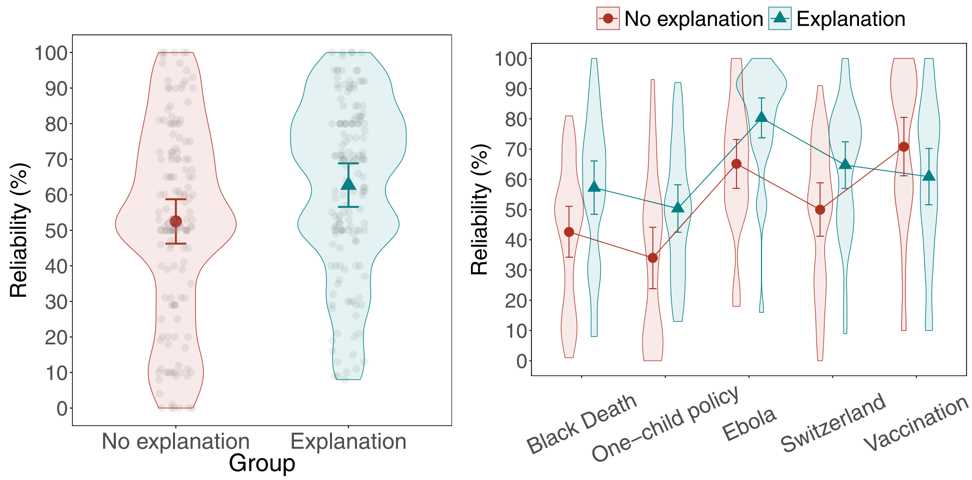
Together, however, results from Experiments 2 and 3 suggest that reliability is mediating the effects of explanation on confidence, further implying that reliability could also have causal effects on confidence. We explore this in the next two experiments.

5. Experiment 4

Experiment 1 showed that explanations can affect confidence and Experiments 2 and 3 further indicated (i) that explanations also have an effect on reliability and (ii) that explanation's effects on confidence are mediated by reliability, suggesting that reliability could causally affect confidence. In this experiment we explore potential causal effects of reliability on confidence. Given the findings in the previous three experiments, we expected that people's confidence estimates will depend on the explainers' reliability levels.

5.1. External expertise and perceived expertise

The method adopted in this experiment to manipulate reliability was through changing the levels of expertise of the explainer: the higher the level of expertise the higher the reliability. However, in the literature one can find multiple notions of expertise. Before we go on to explore the impact of reliability on confidence, it is worth drawing a distinction between at least two kinds of expertise: external expertise



(a) (b)

Fig. 7. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *reliability estimates* in Experiment 3. Gray points are raw data values (jittered along the *x*-axis for visibility) with violin plots showing the frequency of the raw data. (b) The observed data means (with 95% confidence intervals) and violin plots for each scenario broken down for each explanation conditions.

and perceived expertise. External expertise is judged by referring to a person's externally measurable criteria: their qualifications, their track records of success or their experience in a particular activity (see Collins & Evans, 2008). For example, doctors are considered experts according to these external criteria as they have the required qualifications and relevant experience. This kind of expertise has been found to have a significant effect on people's beliefs. For instance, research on the influence of expert testimony on jurors' decision-making suggests that the expert's credentials have a significant effect on jurors' decisions (Krauss & Sales, 2001).

Perceived expertise, on the other hand, is not concerned with the expert's externally measurable criteria. Rather, it has to do with an expert's general demeanor, such as the internal consistency of their remarks (Collins & Evans, 2008). For instance, judges and jurors would perceive an expert's testimony as more believable if it is internally consistent and coherent compared to one that is less coherent, even though the judges and jurors are not themselves domain experts. Zemla et al. (2017) similarly point to the distinction between external expertise and perceived expertise and suggest that external expertise may be mediated by perceived expertise when it comes to the impact of expertise on the goodness of explanations.

In this experiment, we manipulated explainers' external expertise. The external expertise, we believe, would have an effect on participants' perceived expertise of the explainers, which is what is being measured by asking participants to provide their estimates of the reliability of the explainers in the scenarios (see also Zemla et al., 2017). Experiments 2 and 3 suggested that the presence/absence of an explanation for a claim has an impact on the explainer's perceived reliability, and that the explainer's perceived reliability mediates the effects of explanation. We thus expected that the impact of providing an explanation will be attenuated by the external expertise. As we manipulated the level of the external expertise in this experiment, we were able to explore how the effects of explanation on confidence are attenuated by the different levels of external expertise.

5.2. Methods

5.2.1. Participants & design

A total of 161 participants ($N_{FEMALE} = 112$, one participant identified as neither male nor female, $M_{AGE} = 36.5$ years) were recruited from Prolific Academic (www.prolific.co). The selection criteria and the remuneration rate per hour were the same as in the previous experiments. Participants took on average 14.8 min to complete the experiment.

A between-participant design was adopted and participants were randomly allocated in one of 2 (no explanation or explanation) × 2 (reliability: low or high) = 4 groups ($N_{NO\_EXPL\_LOW} = 40$, $N_{NO\_EXPL\_HIGH} = 42$, $N_{EXPL\_LOW} = 40$, $N_{EXPL\_HIGH} = 39$). All participants were asked to provide estimates on two dependent variables: confidence and reliability.

5.2.2. Materials

We used the same 5 scenarios as before, with some further modifications to manipulate the explainer's reliability as either high or low. This was done by introducing the explainer in the preamble of each scenario as either a domain expert (high reliability) or a novice/lay person (low reliability). For example, the *Black Death* scenario read as follows (note that the text in curly brackets appeared only in the explanation condition):

> [*Preamble in the low reliability condition*:] Dave and Jimmy are high school students who are assigned a student project to find out as much as they can on one of the most devastating pandemics in human history, namely the Black Death.
>
> [*Preamble in the high reliability condition*:] Dave and Jimmy are senior researchers at a well-established institute for global health and part of the project investigating devastating pandemics in human history. During a planning meeting they touched upon the Black Death.
>
> [*The rest of the scenario was the same for both the low and high reliability conditions.*]
>
> **Dave**: Millions of people died from the Black Death in the 14th century. I think our project should in part focus on how the Black Death ended.
>
> **Jimmy**: Yes, I agree. Do you already have an idea regarding how the Black Death came to an end?
>
> **Dave**: The Black Death subsided mostly through the use of quarantines.
>
> {**Jimmy**: How so?
>
> **Dave**: People mostly stayed out of the path of infected individuals, rats, and fleas. The uninfected would typically remain in their homes and only leave when it was necessary. Those with the financial resources would traditionally escape to the country, far away from the Black Death-infested cities.}

The expert explainers in other scenarios were: an immunologist (*Vaccination* scenario), an experienced policy-maker who specialized on East Asia (*One-child policy* scenario), a medical practitioner who was part of the Doctors Without Borders team in West Africa treating various epidemic diseases (*Ebola* scenario), and history professors who have been awarded a research grant for a project on armed neutrality in World War II (*Switzerland in WWII* scenario). The non-expert explainers were: a subway operator (*Vaccination* scenario), a person who has just started their undergraduate studies in philosophy (*One-child policy* scenario), a non-medically educated person who read in the news about a team of doctors in West Africa who contracted Ebola (*Ebola* scenario), and a high school student (*Switzerland in WWII* scenario).

5.2.3. Procedure

The procedure was exactly the same as that of Experiment 2, with each participant being asked both the confidence question and the reliability question. Eliciting reliability estimates from the participants provided us with information regarding the success of the reliability manipulation. Additionally, since reliability estimates reflect participants' perceived reliability or expertise of the explainers, they would also provide insight into whether the external expertise manipulation was successful in affecting the participants' perceived reliability or expertise of the explainers.

5.3. Results & discussion

A separate analysis was conducted for each dependent variable.

5.3.1. Confidence

To test the effect of explanation and reliability on people's confidence in statements, we built an LMM with explanation and reliability as fixed effects and a random intercept for each participant. We found a main effect of explanation ($t(157) = -3.28$, $p = .001$) and of reliability ($t(157) = -5.79$, $p < .001$), and no interaction between the fixed effects ($t(157) = -1.1$, $p = .28$). Further, the inclusion of the predictors in the model led to a significant improvement in model fit ($\chi^2(3) = 40.6$, $p < .001$), compared to just having an intercept as a predictor. This suggested that both explanation and reliability have (additive) causal effects on confidence. These results replicate the findings from the previous three experiments and show that providing an explanation does have a significant effect on confidence in a claim's estimates. Specifically, we found that confidence estimates are higher when an explanation is present compared to when no explanation was provided (see Fig. 8(a)). This effect is again replicated across the five scenarios (see Fig. 9).

The main effect of explainer's reliability provides support for the hypothesis that not only presence of an explanation, but also the explainer's reliability have significant effects on confidence estimates. We found that confidence estimates are higher when the explainer's reliability is high than when their reliability is low (Figs. 8(a) and 9).

Fig. 8(a) also shows that the difference between confidence estimates is lower when the explainer's reliability is high compared to when their reliability is low. We found that the difference between the explanation group's confidence estimates ($EMM = 60.5$) and the no explanation group's ones ($EMM = 48.5$) was higher in the low reliability conditions (the absolute difference between means was 12) than in the high reliability conditions where the absolute difference between the explanation group's confidence estimates ($EMM = 73.3$) and the no explanation group's ones ($EMM = 67.3$) was equal to 6. From Fig. 9 we can see that this was also true across the five scenarios. However, as the interaction between the presence of explanation and explainer's reliability was not statistically significant in the above LLM, this finding provides only limited support for the hypothesis that the effects of explanation on confidence are attenuated by different levels of reliability. We further explore this interaction in Experiment 5 below.

5.3.2. Reliability

To test the effect of explanation and reliability on people's reliability estimates, we built an LMM with explanation and reliability as fixed effects and a random intercept for each participant. We found a main overall effect of explanation ($t(157) = -2.03$, $p = .045$) and of reliability ($t(157) = -9.12$, $p < .001$), and no interaction between the fixed effects ($t(157) = -1.39$, $p = .17$). Further, the inclusion of the predictors in the model led to a significant improvement in model fit ($\chi^2(3) = 72.3$, $p < .001$), compared to just having an intercept as a predictor. Similarly to the above findings on confidence there was an effect of explanation and a highly significant effect of the reliability manipulation on participants' reliability estimates suggesting that participants in the high reliability condition provided higher estimates regarding the explainer's reliability compared to those in the low reliability condition. This suggests that the reliability manipulation was successful and that the participants' perceived reliability estimates were affected by the external expertise of the explainer. Figs. 8(b) and 10 show participants' reliability estimates.

6. Experiment 5

In Experiment 4, we found a non-significant interaction between reliability and explanation. Specifically, Experiment 4 suggested that the effect of explanation was attenuated in the high reliability condition compared to the low reliability condition. This finding is consistent with the results from Experiment 2, where we demonstrated that the effect of explanation on confidence is mediated by reliability.

Experiment 4, however, was underpowered to explore this interaction. To achieve sufficient power (80%) to investigate an interaction where a significant proportion (e.g. 70%) of an effect (e.g. presence of explanation) was mediated by a certain factor (e.g. reliability) we would have needed at least 4 times the sample size per cell compared to a well-powered experiment that explores the effect of only one factor (Simonsohn, 2014). This implies that to explore the interaction between explanation and reliability where a significant proportion of the effect of explanation was mediated by reliability as suggested in Experiment 2, we would have required 4 times the number of participants per cell as in Experiment 1, where we only explored the effect of one factor, i.e. explanation. In Experiment 1, we had around 60 participants per cell, which means that in Experiment 4 we would have needed $4 \times 60 = 240$ participants per cell to have enough power to explore the interaction. This required sample size is significantly higher than the number of participants per cell we actually recruited in Experiment 4 (i.e. around 40).

Experiment 5 aims to explore the interaction between the presence/absence of an explanation and the explainer's reliability with a sufficiently higher number of participants to achieve the required power. Experiment 5 uses the same procedure and materials as Experiment 4, and as such it will also serve as a replication of Experiment 4. We aim to recruit around $4 \times 60 = 240$ participants per cell, totaling 960 participants for Experiment 5.

6.1. Methods

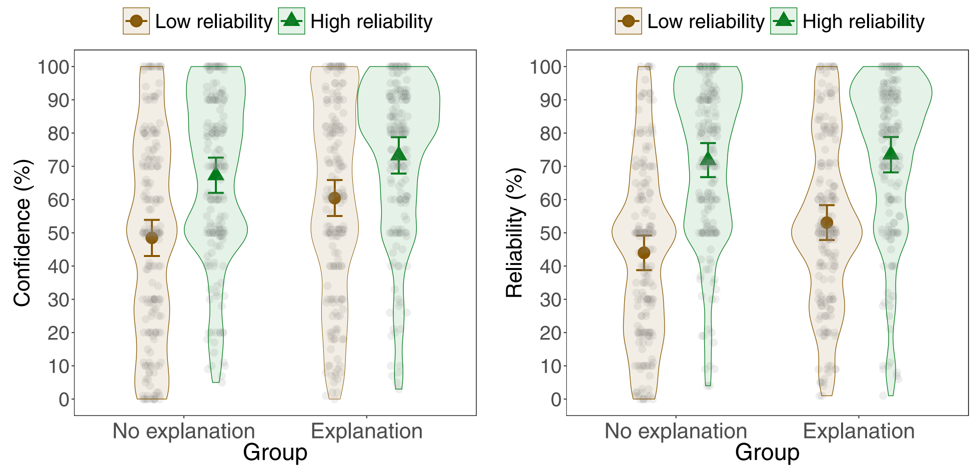
6.1.1. Participants & design

A total of 959 participants ($N_{\text{FEMALE}} = 557$, 7 participants identified as neither male nor female, $M_{\text{AGE}} = 44.1$ years) were recruited from Prolific Academic (www.prolific.co). The selection criteria and the remuneration rate per hour were the same as in the previous experiments. Participants took on average 15.1 min to complete the experiment.

The design was identical to that in Experiment 4. The number of participants in each cell was as follows: $N_{\text{NO\_EXPL\_LOW}} = 241$, $N_{\text{NO\_EXPL\_HIGH}} = 237$, $N_{\text{EXPL\_LOW}} = 240$, $N_{\text{EXPL\_HIGH}} = 241$.

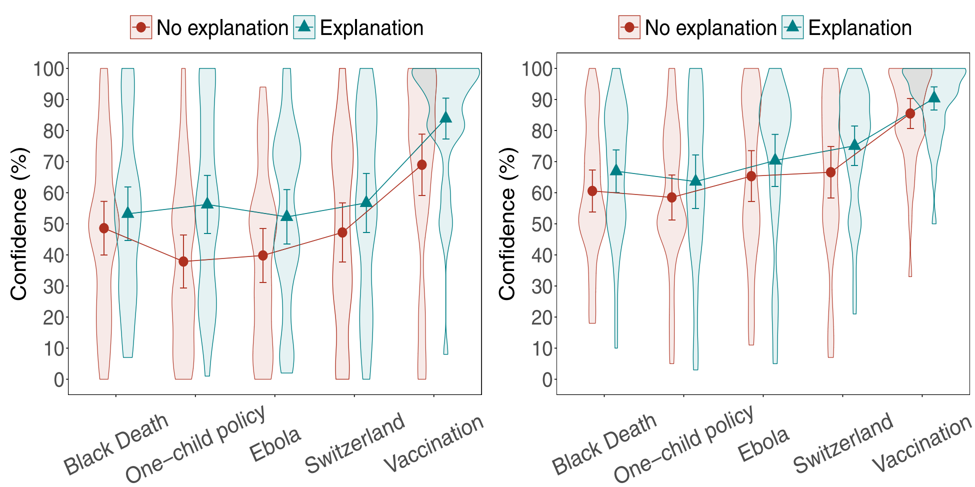
6.1.2. Materials & procedure

Materials and procedure were identical to Experiment 4.



(a) (b)

Fig. 8. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *confidence estimates* in Experiment 4. (b) The EMMs (with 95% confidence intervals) from the LMM built on participants' *reliability estimates* in Experiment 4.



(a) (b)

Fig. 9. (a) The observed data means (with 95% confidence intervals) for participants' *confidence estimates* in each scenario in *the low reliability* condition. (b) The observed data means (with 95% confidence intervals) for participants' *confidence estimates* in each scenario in *the high reliability* condition.

6.2. Results & discussion

A separate analysis was conducted for each dependent variable.

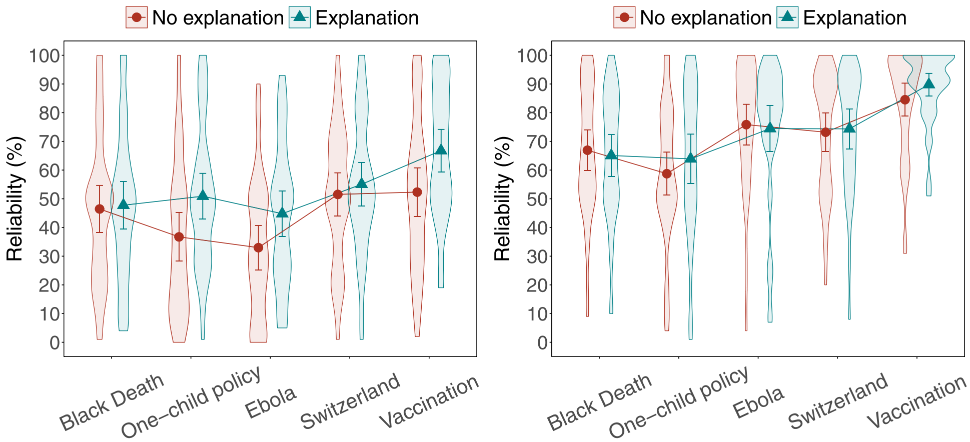
6.2.1. Confidence

Similar to Experiment 4, we built an LMM with explanation and reliability as fixed effects and a random intercept for each participant. We found a main effect of explanation ($t(959) = -7.99$, $p < .001$) and of reliability ($t(959) = -12.44$, $p < .001$), and a significant interaction between the fixed effects ($t(959) = -2.97$, $p = .003$). Further, the inclusion of the predictors for the model led to a significant improvement in model fit ($\chi^2(3) = 204.6$, $p < .001$), compared to just having an intercept as a predictor.

The results replicate those in Experiment 4. We found both the main effects of explanation and reliability. Additionally, however, we also found a significant interaction between explanation and reliability. Specifically, we found that the effect of explanation was attenuated when the explainer's reliability was high compared to when it was low: in the low reliability conditions the absolute mean difference between the explanation group's confidence estimates ($EMM = 60.25$) and the no explanation group's ones ($EMM = 49$) was 11.25 and in the high reliability conditions the difference between the explanation group's confidence estimates ($EMM = 70$) and the no explanation group's ones ($EMM = 64.8$) was 5.2 (see Fig. 11(a)). These results also hold across the five scenarios (see Fig. 12). In other words, the effects of providing an explanation on confidence are most mediated by the reliability of an explainer when that reliability is high; when the reliability of an explainer is low, a significantly smaller amount of the effect of explanation on confidence is mediated.

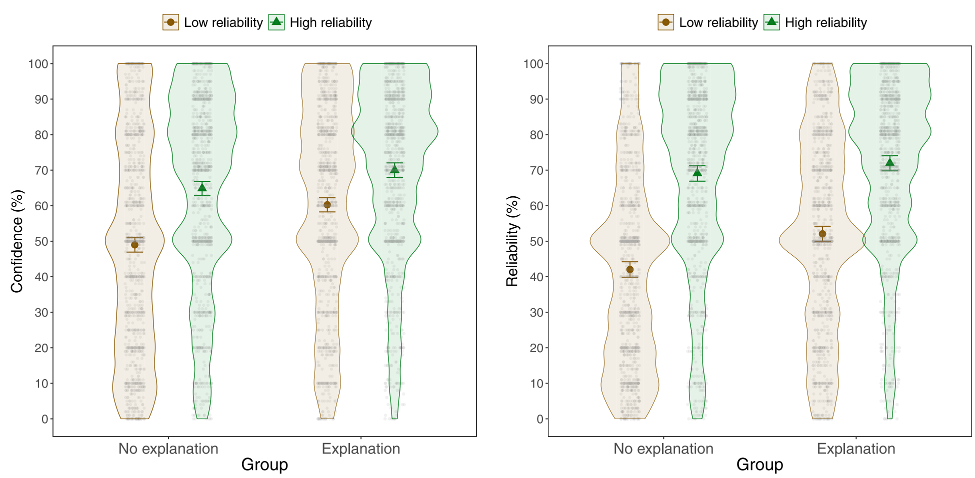
6.2.2. Reliability

To test the effect of explanation and reliability on people's reliability estimates, we built an LMM with explanation and reliability as fixed effects and a random intercept for each participant. We found a main overall effect of explanation ($t(959) = -5.85$, $p < .001$) and of reliability ($t(959) = -21.3$, $p < .001$), and a significant interaction between the fixed effects $t(959) = -3.26$, $p = .001$). Further, the inclusion of the predictors for the model led to a significant improvement in model fit ($\chi^2(3) = 402.2$, $p < .001$), compared to just having an intercept as a predictor. Like in Experiment 4, we found an effect of explanation and a



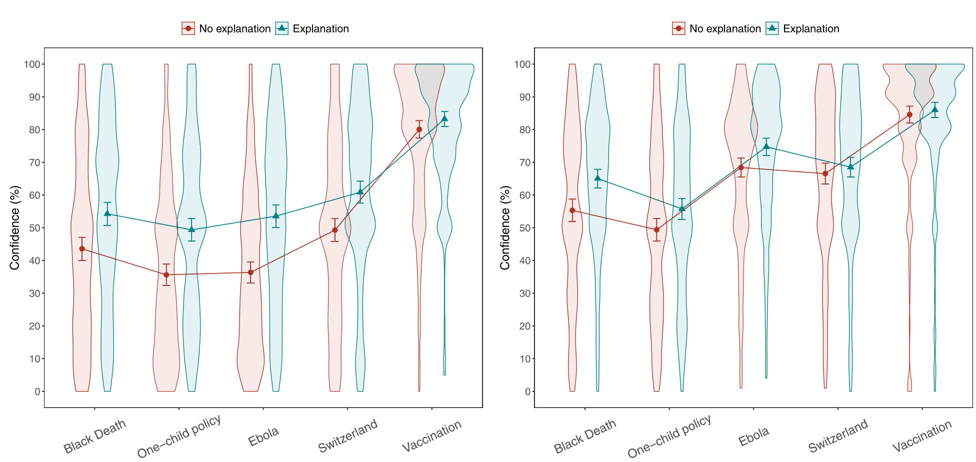
(a) (b)

Fig. 10. (a) The observed data means (with 95% confidence intervals) for participants' *reliability estimates* in each scenario in *the low reliability* condition. (b) The observed data means (with 95% confidence intervals) for participants' *reliability estimates* in each scenario in *the high reliability* condition.



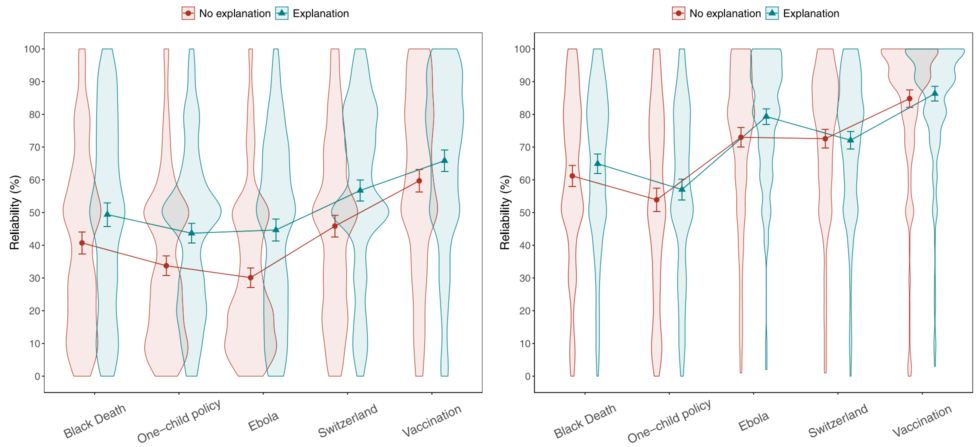
(a) (b)

Fig. 11. (a) The estimated marginal means (with 95% confidence intervals) from the LMM built on participants' *confidence estimates* in Experiment 5. (b) The EMMs (with 95% confidence intervals) from the LMM built on participants' *reliability estimates* in Experiment 5.



(a) (b)

Fig. 12. (a) The observed data means (with 95% confidence intervals) for participants' *confidence estimates* in each scenario in *the low reliability* condition. (b) The observed data means (with 95% confidence intervals) for participants' *confidence estimates* in each scenario in *the high reliability* condition.



(a)

(b)

Fig. 13. (a) The observed data means (with 95% confidence intervals) for participants' *reliability estimates* in each scenario in *the low reliability* condition. (b) The observed data means (with 95% confidence intervals) for participants' *reliability estimates* in each scenario in *the high reliability* condition.

highly significant effect of the reliability manipulation on participants' reliability estimates suggesting a successful reliability manipulation. In addition, we also found a significant interaction effect. This suggests that the effect of external reliability on perceived reliability may be attenuated by explanation (see Figs. 11(b) and 13). This is an interesting finding as it implies that not just confidence, but also perceived reliability may be affected by both the presence of an explanation and the explainer's external reliability.

7. General discussion

7.1. Summary of the main findings

We carried out five experiments to explore the relationship between explanation, reliability, and confidence in claims. Experiment 1 provided evidence that explanations have a significant impact on our confidence in a claim, consistent with previous literature on the effects of explanations.

Experiments 2 and 3 introduced an aspect of explanation, namely reliability, that becomes apparent only when explanation is fully considered as a social act of communication. The results from these experiments revealed that a (good) explanation not only increased participants' confidence in a claim (replicating the findings from Experiment 1) but also heightened their perceived reliability of the explainer. Furthermore, the findings suggested that much of the effect of providing an explanation on confidence is mediated by the explainer's reliability, indicating that the reliability of the explainer may have a causal influence on confidence in a claim.

Experiment 4 further delved into the potential causal effect of reliability on confidence, while Experiment 5 explored the interaction between explanation and reliability. The results from these experiments provided several key insights: (i) the explainer's reliability significantly influences confidence in a claim, (ii) the effect of providing a good explanation on confidence is more pronounced when the explainer's reliability is low compared to when it is high, suggesting that the impact of explanation on confidence is largely mediated by reliability in high-reliability conditions, (iii) participants' reliability estimates were mainly guided by the level of expertise exhibited by the explainer, and (iv) the presence of an explanation had different effects on participants' confidence in the claim when the explainer's reliability was high compared to when it was low.

7.2. Relations to extant research

These findings not only complement but also extend previous research on explanation. For example, Koehler (1991) conducted experiments where participants were asked to provide explanations of hypotheses, and they found that explanations boost confidence in those hypotheses. In the five experiments conducted here, we closely replicated these findings using a different paradigm, where participants were given explicit explanations without being asked to produce them. Additionally, these findings extend onto everyday explanations of real-world events. The work on everyday explanations is still very much in its infancy and often limited to correlational analyses of different explanatory criteria (Zemla et al., 2017). Experiments presented here went beyond mere correlational analyses and instead employed randomized allocation of participants to different conditions, contributing to a deeper understanding of the impacts of everyday explanations on our beliefs.

This study has also delved into the social aspects of explanation, particularly by manipulating and studying the reliability of the explainer. It sheds light on the interplay between providing an explanation and the reliability of the explainer, especially in the context of everyday explanations. Notably, the results suggest that in certain cases, a non-expert providing a (good) explanation can have a similar impact on our beliefs as an expert making a claim without providing an explanation. These findings align with predictions from formal models of source reliability, such as the Bovens and Hartmann (BH) model (Bovens & Hartmann, 2003). According to the BH model, higher reliability of a source leads to a greater impact of evidence on the confidence in a hypothesis. The findings from Experiments 4 and 5 suggested that this is the case in both the no explanation and the explanation condition: we found a significant effect of reliability on the claim that is suggesting how data/evidence may have come about (e.g. how the Black death ended). Furthermore, if we slightly modify the BH model to include explanation as an additional variable whose effect is (partially) mediated by the reliability of the explainer, then one would expect that in such a model the impact of explanation is higher when the reliability is low compared to when the reliability is high. One can also explore parametarizations of that model where the impact of explanations when the reliability is low is similar to the impact of a high reliability source that is not providing an explanation. These are, however, only (plausible) conjectures and call for further exploration.

The effects of reliability on people's beliefs, though underexplored, could potentially account for some of the findings in the literature on the effects of explanations. For example, Weisberg, Keil, Goodstein, Rawson, and Gray (2008) showed that adding irrelevant neuroscience information had a particularly striking effect on non-experts' judgments of bad explanations. Specifically, non-experts judged these explanations significantly more satisfying than the bad explanations without the irrelevant neuroscience information.[3] Furthermore, the results of Experiment 2 from Weisberg et al. (2008) suggested that there was no significant difference between the satisfaction ratings of the good explanations without any irrelevant neuroscience information and the bad explanations with this information. Weisberg, Taylor, and Hopkins (2015) provide a few potential explanations for these results. However, another explanation that they have not considered comes from the findings of our five experiments. Namely, it is plausible that the presence of the irrelevant neuroscience information indicated to the participants that the explanation is coming from a reputable and reliable source with expertise in neuroscience. Drawing from the findings from Experiments 4 and 5, one could expect that the participants' rating of the bad explanations would be higher when such information is present, potentially even having similar ratings as the good explanations that lack such information.

7.3. Prior knowledge

Throughout the paper, we have observed indications of the effects of prior or background knowledge on both confidence in claims and the reliability of an explainer. In certain scenarios (e.g. *Vaccination* and *Ebola* scenarios), participants' confidence and reliability estimates in both the no explanation and explanation conditions were high and the difference in estimates between the conditions was small. Textual responses from participants in these scenarios indicated that they provided high estimates for confidence and reliability questions because the claims and explanations aligned with their prior knowledge. This suggests that the effects of explanation and reliability may be attenuated when prior knowledge is significant.

This observation is supported by existing research. Anderson and Sechler (1986) and Koehler (1991) suggest that individuals who already hold a strong opinion about a topic are unlikely to be influenced by an explanation. On the other hand, individuals who do not have a strong opinion or do not possess prior knowledge about a topic are more susceptible to the effects of explanation. Specifically, the impact of explanation tends to be lower when people are already familiar with the topic in question. This could be because people are less likely to seek an explanation for something they are already familiar with (Lombrozo, 2012) and/or because they already hold a high level of confidence in the claim (Thagard, 1989). Koslowski (1996, Experiments 9 and 10) found that the presence of a mechanism (explanation) had a larger effect on factors (causes) perceived as implausible compared to those perceived as plausible. Similar effects of prior knowledge on explanation have been observed in category learning (Williams & Lombrozo, 2013). Regarding the relationship between prior knowledge and reliability, Jarvstad and Hahn (2011, Experiment 2) suggested that participants' prior confidence in the claims affected their judgments of source reliability. A source was judged as more reliable if it provided a statement in which participants had higher prior confidence.

Further studies are needed to explore the effects of prior knowledge in the context of explanation and reliability in more detail. However, existing research and the studies presented in this paper suggest that these effects are not negligible. Considering the impact of prior knowledge is particularly important in studies that employ everyday explanations, as participants are more likely to be familiar with the claims and explanations presented to them.

[3] Similar results were found when non-experts were asked to judge the quality of research based on the abstracts of papers. Eriksson (2012) found that non-experts judged the research to be of higher quality if it included equations, even though the equations did not make sense in the context of the research.

7.4. Future work

As mentioned in the previous section, prior knowledge can have significant effects on our confidence and reliability estimates. Although we did not quantitatively measure or control for this effect in our study, participants' textual responses suggested that prior knowledge did influence our experiments. This might have led to a reduced effect of explanation on confidence and reliability. Despite this, we observed clear effects of explanation and reliability, especially in scenarios where participants did not report prior familiarity with the topic (e.g. *One-child policy* and *Switzerland in WWII*). Future research should explicitly consider and explore the effects of prior knowledge when investigating the relationships among explanation, reliability, and confidence.

In the experiments presented in this paper, we aimed to manipulate the perceived reliability levels of the explainers by communicating the different levels of external expertise. The results from Experiments 4 and 5 suggest that this manipulation was successful. Another potential way of manipulating perceived reliability could be through including technical information relevant to the domain, as suggested by the findings of Weisberg et al. (2008). However, neither of these approaches directly manipulates reliability; they both manipulate some other factor (external expertise or the presence of technical information) to affect the (perceived) reliability of the explainer. Hahn et al. (2009) propose a potentially more direct way to manipulate reliability by communicating to participants that the information comes from a reliable or unreliable source. These and other methods for manipulating the reliability of the source should be investigated in future research on explanations.

In the studies presented in this paper, we only considered the impact of the presence/absence of "good" explanations. However, it is plausible that the presence of poor explanations may not increase confidence in a claim, but rather decrease it. Additionally, Zemla et al. (2017) suggest that perceived expertise (reliability) mediates the effect of external expertise (reliability) on the perceived goodness of explanations. Moreover, findings from Experiment 5 showed that both the presence of an explanation and explainer's reliability impact perceived reliability, and there is an interaction between the two. It is reasonable to expect that the quality of explanations also affects assessments of reliability. As a result, poor explanations might negatively impact explainer's perceived reliability and interact with the presence/absence of such explanations, producing interesting (interaction) effects on confidence in a claim. This should be explored in further studies.

7.5. Conclusion

Explanations play an important role in both intra-personal and inter-personal reasoning and decision-making. Considering the social aspects of explanations is essential for a comprehensive understanding of their role. In this paper, we have explored the effects of the reliability of an explainer, which is just one of the social aspects of explanations. Further research should take into account not only explainer reliability but also other characteristics of explanations understood as communicative acts.

CRediT authorship contribution statement

Marko Tešić: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Ulrike Hahn:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Data availability

All data and materials are publicly available via the OSF at https://osf.io/wkhdc/.

References

Ahn, W.-k., Novick, L. R., & Kim, N. S. (2003). Understanding behavior makes it more normal. *Psychonomic Bulletin & Review*, *10*(3), 746–752.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: the role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, *39*(6), 1037.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology*, *50*(1), 24.

Antaki, C., & Leudar, I. (1992). Explaining in conversation: Towards an argument model. *European Journal of Social Psychology*, *22*(2), 181–194.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Bechlivanidis, C., Lagnado, D., Zemla, J. C., & Sloman, S. (2017). Concreteness and abstraction in everyday explanation. *Psychonomic Bulletin & Review*, *24*(5), 1451–1464.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford University Press.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science*, *24*(4), 573–604.

Collins, H., & Evans, R. (2008). *Rethinking expertise*. University of Chicago Press.

Collins, P., & Hahn, U. (2019). We might be wrong, but we think that hedging doesn't protect your reputation. *Journal of experimental psychology. Learning, memory, and cognition*.

Collins, P., Hahn, U., von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in Psychology*, *9*, 18.

Davoodi, T., & Lombrozo, T. (2022). Varieties of ignorance: Mystery and the unknown in science and religion. *Cognitive Science*, *46*(4), Article e13129.

Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. *The Philosophical Quarterly*, *63*(252), 428–444.

Douven, I. (2017). Abduction. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University, https://plato.stanford.edu/archives/sum2017/entries/abduction/.

Douven, I., & Schupbach, J. N. (2015). The role of explanatory considerations in updating. *Cognition*, *142*, 299–311.

Eriksson, K. (2012). The nonsense math effect. *Judgment and Decision Making*, *7*(6), 746.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology*, *75*(1), 105–118.

Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science*, *6*(2), 172–182.

Hahn, U., Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*(4), 337–367.

Hahn, U., Harris, A. J., & Corner, A. (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in Cognitive Science*, *8*(1), 180–195.

Hahn, U., Oaksford, M., & Harris, A. J. (2013). Testimony and argument: A bayesian perspective. In *Bayesian argumentation* (pp. 15–38). Springer.

Harman, G. (1965). The inference to the best explanation. *The Philosophical Review*, *74*(1), 88–95.

Harris, A. J., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, *40*(6), 1496–1533.

Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*(2), 135–175.

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*(1), 65.

Holden, C. (2001). Music for relactation. *Science*, *293*(5528), 205.

Jarvstad, A., & Hahn, U. (2011). Source reliability and the conjunction fallacy. *Cognitive Science*, *35*(4), 682–711.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*(1), 601–625.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254.

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*(2), 259–300.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher, & W. Salmon (Eds.), *Scientific explanation: Minnesota studies in the philosophy of science* (pp. 410–505). Minneapolis: University of Minnesota Press, Minneapolis.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, *110*(3), 499.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 461.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Mit Press.

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, *7*(2), 267.

Lagnado, D. (1994). *The psychology of explanation: A bayesian approach* (Unpublished Masters thesis), Schools of Psychology and Computer Science, University of Birmingham, UK.

Lipton, P. (2003). *Inference to the best explanation*. Routledge.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*(3), 232–257.

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

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*(10), 748–759.

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2018). Partial source dependence and reliability revision: The impact of shared backgrounds. In *Proceedings of the 40th annual conference of the cognitive science society*.

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source dependence on belief and reliability revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1795–1805.

Merdes, C., Von Sydow, M., & Hahn, U. (2020). Formal models of source reliability. *Synthese*, 1–29.

Olsson, E. J., & Vallinder, A. (2013). Norms of assertion and communication in social networks. *Synthese*, *190*(13), 2557–2571.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*(1–2), 123–163.

Psillos, S. (2000). Abduction: Between conceptual richness and computational complexity. In *Abduction and induction* (pp. 59–74). Springer.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429.

Ross, L. D., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology*, *35*(11), 817.

Sherman, S. J., Zehner, K. S., Johnson, J., & Hirt, E. R. (1983). Social explanation: The role of timing, set, and recall on subjective likelihood estimates. *Journal of Personality and Social Psychology*, *44*(6), 1127.

Simonsohn, U. (2014). No-way interactions. http://dx.doi.org/10.15200/winn.142559.90552.

Singmann, H., & Kellen, D. (2019). An introduction to linear mixed modeling in experimental psychology. In *New methods in cognitive psychology* (pp. 4–31). Psychology Press.

Thagard, P. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*(2), 76–92.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38.

Toulmin, S. E. (1958/2003). *The uses of argument*. Cambridge University Press.

Trout, J. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, *69*(2), 212–233.

Trout, J. (2008). Seduction without cause: Uncovering explanatory neurophilia. *Trends in Cognitive Sciences*, *12*(8), 281–282.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

Walton, D. (2004a). A new dialectical theory of explanation. *Philosophical Explorations*, *7*(1), 71–89.

Walton, D. (2004b). *Relevance in argumentation*. Routledge.

Walton, D. (2007). *Witness testimony evidence: Argumentation and the law*. Cambridge University Press.

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.

Weisberg, D. S., Taylor, J. C., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgment and Decision Making*, *10*(5), 429.

Wilkenfeld, D. A., Plunkett, D., & Lombrozo, T. (2016). Depth and deference: When and why we attribute understanding. *Philosophical Studies*, *173*(2), 373–393.

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*(1), 55–84.

Zemla, J. C., Sloman, S., Bechlivanidis, C., & Lagnado, D. (2017). Evaluating everyday explanations. *Psychonomic Bulletin & Review*, *24*(5), 1488–1500.