Please quote the published version

Gaszczyk, G., Krogulska, A. Lying by explaining: an experimental study. *Synthese* 203, 80 (2024). https://doi.org/10.1007/s11229-023-04451-5

Lying by explaining: An experimental study

Abstract

The widely accepted view states that an intention to deceive is not necessary for lying. Proponents of this view, the so-called non-deceptionists, argue that lies are simply insincere assertions. We conducted three experimental studies with false explanations, the results of which put some pressure on non-deceptionist analyses. We present cases of explanations that one knows are false and compare them with analogical explanations that differ only in having a deceptive intention. The results show that lay people distinguish between such false explanations and to a higher degree classify as lies those explanations that are made with the intention to deceive. Non-deceptionists fail to distinguish between such cases and wrongly classify both as lies. This novel empirical finding indicates the need for supplementing non-deceptionist definitions of lying, at least in some cases, with an additional condition, such as an intention to deceive.

1 Introduction

Sometimes we want to explain a phenomenon that is too difficult to understand for our audience. In such a case we can explain the phenomenon fully—being aware that the audience probably would not understand it, not explain anything, or we can explain the material in a simplified way, often by knowingly presenting strictly false information. An example of the latter communicative practice are explanations performed by teachers. Walsh and Currie (2015, 424), for instance, say that "*The truth, the whole truth and nothing but the truth* is no teacher's maxim." The use of false information in educational contexts is considered to be justifiable or even judged as praiseworthy. Elgin (2007) calls such explanations felicitous falsehoods. At the same time, however, such explanations have been labelled as caricatures (Walsh and Currie 2015), disinformation (Fallis 2015), and even lies. Consider the reasoning behind the last label: "A lie-to-children is a statement that is false, but which nevertheless leads the child's mind towards a more accurate explanation, one that the child will only be able to appreciate if it has been primed with the lie" (Pratchett et al. 1999, 38).

An example of such a lie-to-children can be found in teaching physics, where the Bohr model of an atom—which depicts an atom like a little solar system where electrons travel in circular orbits around the nucleus—is still often taught as representing the atomic structure. We know that atoms do not behave as the Bohr model predicts; nevertheless, this model is still used before more complex models are introduced. Such examples are not rare and peripheral—their usage is widespread, especially when it comes to explicating scientific phenomena.[1] Thus, an explanation of something false is often done with the promise that it will be corrected in the future. Independently of this promise, this example shows a crucial aspect of explanation, i.e., a good explanation is directed towards and adjusted to the audience. The teacher in a primary school cannot explain what science nowadays says about the structure of an atom—such an explanation would be incomprehensible to the audience due to its complexity.

Are teachers who explain such theories as the Bohr model lying? On the one hand, they follow a curriculum and widespread educational practice; moreover, they do that to spread understanding in their pupils about the structure of an atom, even if it is only a partial understanding. On the other hand, however, they knowingly say something they believe to be false, which by most definitions of lying is sufficient to lie. More precisely, the dominant view nowadays is that lies are insincere assertions.[2] The consensus is that an intention to deceive is not necessary for lying. Proponents of such a view are called non-deceptionists. Their opponents, the deceptionists, claim that an intention to deceive is a necessary condition in a definition of lying.[3] The status of an intention to deceive has been a subject of empirical investigation

[1] For examples in the philosophy of science, see e.g. Elgin 2007, 2017; de Regt and Gijsbers 2017; de Regt 2017; in epistemology, see e.g. McKinnon 2015; Goldberg 2015; in humanities, see Walsh and Currie 2015.

[2] See e.g. Carson 2006; Sorensen 2007; Fallis 2009; Saul 2012; Stokke 2018; Marsili 2020, Viebahn 2021.

[3] See e.g. Isenberg 1964; Primoratz 1984; Williams 2002; Mahon 2008, 2016; Lackey 2013; Meibauer 2014a.

(see Section 2.3). However, the debate has been primarily focused on specific cases—the so-called bald-faced lies. Crucially, false explanations have not been empirically tested yet—we aim to bridge this gap. Whether one lies in such cases is an empirical matter and thus we empirically investigate intuitions regarding such explanations.

The aim of this paper is to empirically test whether the speaker making a false explanation, such as the explanation of the Bohr model, is considered by ordinary people to be lying. The central question of our study is the following one: "Does modification of the speaker's intention (positive/negative) change the lie attribution?" We want to test whether ordinary speakers agree that the speaker with a negative intention (for instance, wants to mislead the audience) lies while the speaker with a positive intention (for instance, presents a simplified material because she believes that presenting a more difficult one would be confusing for the audience) does not lie. Our prediction is that changing the speaker's intention influences the lie attribution: the speaker with the negative intention will be classified as someone who lies to a significantly higher degree than the speaker with the positive intention. Such a result puts some pressure on non-deceptionist definitions of lying, according to which both speakers lie.

The plan is as follows. In Section 2, we present a theoretical part of our work. We make a case for appropriate albeit false explanations and argue that such explanations are not lies. Sections 3-5 report three experiments that focus on empirically testing the relevance of the speaker's intention on lie attributions in false explanations. In Section 6, we discuss the results, focusing on their consequences for non-deceptionists.

2 Explanations and lies

2.1 What is an explanation?

The link between explanation and understanding is widely recognized. One plausible way of expressing this relation is that explanations provide understanding (Lipton 2004; Grimm 2010). In this section, we argue that some false explanations—such as teaching about the structure of an atom by means of simplifications like the Bohr model—can generate understanding and be considered proper in the presented context.

The notion of understanding can be explicated in many ways; the widespread tradition treats understanding as involving an act of grasping.[4] One take is, thus, to think about understanding as "… something like grasping systematic connections among elements of a complex whole, or gaining insight into certain relations between items within a larger body of information" (Jäger 2016, 180). Among various types of understanding, we focus on the so-called objectual type, which concerns subject matters or domains of things (Kvanvig 2003; cf. Wilkenfeld 2013; Kelp 2015; Baumberger and Brun 2017). Furthermore, because our focus will be on scientific explanations, we restrict attention to cases of understanding empirical phenomena, like understanding the structure of the solar system, the phenomenon of evolution, climate change, etc.

Having said that, consider the following case of explanation:

> *Atom Story (positive intention)*
> John is a physics teacher at a summer camp for primary school students. One day, he teaches about the structure of an atom. He has some freedom in deciding how he will explain this topic. He explains the Bohr model of an atom, according to which electrons travel in circular orbits around the nucleus. John knows that this is a crude simplification and a false depiction of an atom. He is aware that there are other, more exact but also much more complicated models of an atom. He presents the Bohr model because he thinks that knowing this model is sufficient on this level of education. He thinks that explaining a more complicated model would only confuse his pupils. As a result of the presentation, his pupils acquire some understanding of the structure of an atom.

Gaszczyk (2023), proposing a detailed analysis of such cases, argues for a normative account of the speech act of explanation with understanding as its norm. Here we

[4] See e.g. Kvanvig 2003; Elgin 2004, 2007, 2017; de Regt and Dieks 2005; Grimm 2006, 2016; de Regt 2009; Khalifa 2013; Strevens 2013; Kelp 2015; Hills 2016.

only sketch a plausible take. It is natural to assume that the speaker should understand the subject of explanation, as John does here. Moreover, a good explainer must take the epistemic position of their audience into consideration and adjust their explanation to the audience's background knowledge and to their capabilities of comprehension. John is aware of the complexity of the phenomenon he wants to explain and for that reason gives an explanation that can be grasped by the audience. Finally, in this context, delivering a more exact explanation would be improper—if John were to do that, he would fail as a teacher since his explanation would not be understood by his audience.

Some could wonder how John's pupils can acquire an understanding if his explanation is (at least) partially false. Factivists about understanding would deny that there is any transfer of understanding in such a case.[5] However, such an explanation is compatible with the non-factive notion of understanding.[6] Non-factivism, among other things, explains the fact that understanding is a gradable notion (Kvanvig 2003; Elgin 2007; Hills 2016; Khalifa 2017). This correlates with our linguistic practice of attributing understanding—one can understand a particular phenomenon barely or partially, but later, one's understanding can be improved, and thus, one can understand the same phenomenon fully or completely.

Following non-factivism, we can say that John's explanation provides at least some genuine understanding. The falsehood that electrons travel in circular orbits around the nucleus allows John's students to grasp the basic concepts of the structure of an atom and the relationship between them. This is corroborated by McKagan et al.'s (2008) empirical research that concerns teaching the Bohr model. They show that starting education from false but simpler theories may help later in a better understanding of more exact theories. Thinking about such cases, Elgin (2009, 325) observes that "… the pattern exhibited in this case is endemic to scientific education. We typically begin with rough characterizations that properly orient us toward the phenomenon, and then refine the characterizations as our understanding

[5] See e.g. Grimm 2006; Khalifa 2012; Strevens 2013; Greco 2014; Kelp 2015; for an overview, see e.g. Baumberger et al. 2017.

[6] See e.g. Zagzebski 2001; Elgin 2004, 2007, 2009, 2017; De Regt and Dieks 2005; De Regt 2009, 2017; De Regt and Gijsbers 2017; Potochnik 2017.

of the science advances."[7] Thus, by his explanation, John lays the ground for expanding the understanding in the future.

2.2 Lying by explaining

Having established that some false explanations can be considered proper, we turn now to the question concerning the sincerity of such explanations. In this section, we argue that John is insincere, but he is not lying. We motivate this view and put it against the dominant approach to lying.

Let us start with the basic account of insincerity. One influential approach to insincerity proposes that each speech act expresses a distinctive propositional attitude (Searle 1969; Bach and Harnish 1979). By expressing an attitude, a speaker represents oneself in a certain way. For instance, an assertion expresses a belief, and thus by asserting that *p* a speaker represents oneself as believing that *p*. Crucially, one can express a particular attitude without having that attitude. By extension, if one asserts that *p* while believing that not *p*, one misrepresents oneself. By doing that, the speaker is insincere. After all, they violate Grice's first sub-maxim of quality (Grice 1989, 27), i.e., they say something they believe to be false.[8]

Explanations are standardly treated as assertions.[9] Following this view, by explaining *p* one represents oneself as believing *p*. Since this standard of sincerity is most widespread, we will also use it throughout the paper.[10]

---

[7] See also Khalifa (2017) who makes a similar observation. However, the crucial difference between non-factivists, like Elgin, and factivists, like Khalifa, is that only the latter maintains that "even minimal understanding requires approximately true explanations" (2017, 21).

[8] Such an account of insincerity is sufficient for our purpose; for various refinements, see e.g. Marsili 2014; Stokke 2014.

[9] See e.g. Turri 2015c; Goldberg 2015; McKinnon 2015; Kelp and Simion 2021. There is also a dissenting view such that explanations are distinct speech act types, see e.g. Achinstein 1983; Gaszczyk 2023. Following the latter view, some may argue that if explanations are not assertions, they cannot be lies. In response, it has been argued that lies are not restricted to assertions (for references, see footnote 16), and thus explanations may also be lie-prone speech acts. Moreover, if John in *Atom Story (negative intention)* is not lying, he is also not lying in proximal cases that are clear instances of lies (like *Atom Story (negative intention)* which we will introduce shortly, cf. footnote 15).

[10] Explanations can be also treated as expressions of understanding. Thus, the idea is that in explaining a particular phenomenon, one represents oneself as understanding the phenomenon in

Let us return to John. He explains something he believes to be false, and because of that, he is insincere. His good intentions or the fact that he follows the curriculum does not change this. Nevertheless, intuitions regarding such cases vary in two opposite directions, i.e., some argue that he is not insincere, and some that he is not only insincere but that he lies. The rest of this section is devoted to arguing against these proposals.

One strategy that is supposed to show that John is not insincere rests on the premise that his explanation is somehow hedged. Gaszczyk (2023) discusses several variants of this strategy. According to one of them, when John in the classroom says *p* (for instance, "Electrons travel in circular orbits around the nucleus"), he uses a covert hedge and really means something like "According to the Bohr model, *p*," or "According to the textbook, *p*."[11] Thus, John is not making a flat-out assertion that *p* and thus he does not commit to the truth of *p*, but rather to the hedged claim that he knows is true. Following this proposal, John's explanation is sincere because he says something he believes to be true.

There are, however, several problems with this proposal. Firstly, it does not deliver any criteria for determining when John explains the hedged content and when he speaks for himself. Consider that in the presented story it is explicitly stipulated that John has some freedom in deciding how he will introduce the material and how deep into the topic he will dive. It is hard to maintain that John is merely reporting what the textbook says since there is no curriculum or textbook to follow—he gives an explanation that he himself decides is proper for his audience. Secondly, there is no criterion for what kind of hedging John is supposed to use. Since there are many possibilities that differ significantly from each other (e.g., "According to one atomic model, *p*," "According to the textbook that I prefer, *p*," or "According to my experience as a teacher in how to introduce this topic, *p*"), this is a major deficiency of this strategy. Finally, even if we would agree that John covertly hedges his explanation, his audience does not know that. Consider that the standard

question. Inversely, if one explains something without understanding it, one can be judged as insincere. For a discussion, see e.g. Gaszczyk 2023.

[11] Milić (2017) pursues this strategy in arguing for the propriety of selfless assertions. We extend it to the present cases; for a detailed critique of this strategy, see Gaszczyk 2019.

questions that they can ask, such as "How do you know that?," or "What makes you believe that?," show that they presuppose that he believes in what he says. This indicates that he is in the context in which the Gricean maxims, particularly the maxim of quality, are in force. After being challenged, he can say that he was covertly hedging his explanation, but this would rather generate confusion and create a sense of insincerity. Thus, we do not find the proposal that John covertly hedged his explanation to be promising.[12]

Others argue that John is not only insincere but that he is lying. This conclusion can be derived from non-deceptionist definitions of lying. All of them are assertion-based definitions, i.e., argue that insincerely asserting is sufficient for lying; they differ in how they define the notion of assertion. We briefly consider two types of such definitions.

The commitment-based definitions of lying (Marsili 2020; Viebahn 2020, 2021; cf. García-Carpintero 2021) maintain that one lies only if one undertakes a proper commitment. For instance, by committing to *p* one is responsible to defend *p* if challenged (cf. Carson 2006, 2010; Saul 2012)—John is responsible for doing that. Viebahn (2020, 2021) emphasises the fact that liars do not retain deniability, i.e., when accused of lying about *p*, liars cannot sincerely and consistently deny saying that *p*. Consider John: when accused of lying, he cannot sincerely and consistently deny that he ever said that electrons travel in circular orbits.[13]

---

[12] This proposal would be more probable (although problematic for the same reasons) if John would deliver his explanation as a teacher in a classroom—some could argue that in such a context he would not speak for himself but merely report what the textbook or educational community tells him to teach. Our story is made in a different context to mitigate such a critique.

[13] For a critique of Viebahn's view, see e.g., Marsili and Löhr 2022; Pepp 2022. However, some commitment-based definitions of lying may be able to accommodate the cases. For instance, Marsili (2020) argues that assertoric commitment has two components. The first is accountability, which refers to "the speaker's *prima facie* liability to be criticised if what they said turns out to be false" (2020, 15). The second is discursive responsibility—in the context of rational discourse, the speaker is expected to perform certain conversational moves, like defending the stated claim if appropriately challenged. As for the former, it may seem that John is not *prima facie* liable to be criticised. However, when it comes to the latter, John is responsible for what he says, e.g., he should defend it when challenged. Still, even if Marsili's definition did not classify false explanations as lies, this would be an exception among

Another view characterises lying in terms of updating the common ground (cf. Stalnaker 1978). Stokke (2018) argues that one lies only if one proposes to make something common ground while believing that it is false. Looking at John, his explanation of the Bohr model counts as a proposal to add this information to the common ground; importantly, he believes that it is false.

Thus, most non-deceptionists agree that John is lying.[14] One strategy to resist this conclusion would be to deny that John is making an assertion in a relevant for lying sense. We think that any such attempts would be unsatisfactory. Any definition of lying that would deny that John makes an assertion would be too narrow and thus wrongly exclude many intuitive cases of lies. For instance, consider this variation of *Atom Story*:

> *Atom Story (negative intention)*
> John is a physics teacher at a summer camp for primary school students. One day, he teaches about the structure of an atom. He has some freedom in deciding how he will explain this topic. He explains the Bohr model of an atom, according to which electrons travel in circular orbits around the nucleus. John knows that this is a crude simplification and a false depiction of an atom. He is aware that there are other, more exact but also much more complicated models of an atom. He presents the Bohr model because he wants to deceive his pupils about the real structure of an atom. He is also aware that explaining a more complicated model would demand from him much more work. As a result of the presentation, his pupils acquire some understanding of the structure of an atom.

---

non-deceptionists. Moreover, if this were the case, it would exclude all false explanations from being lies, which is also unsatisfactory (see the next example).

[14] It seems that other non-deceptionist definitions of lying deliver the same results (see e.g. Sorensen 2007; Fallis 2012, 2013), but a careful examination of each theory is needed to corroborate this observation. With this caveat in mind, in the rest of the paper, for simplicity, we will be saying that non-deceptionists classify explanations like John's as lies.

This is a clear case of lying.[15] The only difference with the previous version of the story is in John's intention—here, he aims to deceive his audience. If someone claims that John is not making an assertion in the first case, it follows that he is also not lying here. This is not plausible. Non-deceptionists classify both versions of *Atom Story* as lies. However, intuitively at least, there is a striking difference between them. We argue that a proper definition of lying should distinguish between these cases—with only the latter one being classified as a lie.

2.3 Previous studies

There is a general agreement that a proper definition of lying should track how this term is used by ordinary language speakers (e.g. Carson 2006; Fallis 2009). Recent experimental studies have been testing ordinary speakers' intuitions regarding various features of lying; such as whether an intention to deceive or actual falsity is necessary for lying, or whether we can lie with speech acts beyond assertions.[16] The results of these experiments influenced how lying is nowadays defined.

Non-deceptionist definitions of lying have been greatly impacted by cases of the so-called bald-faced lies, i.e., lies arguably made without an intention to deceive (e.g. Sorensen 2007; Carson 2010; Fallis 2009, 2013, 2015; Saul 2012; Stokke 2018; cf. Krstić 2019). The experimental studies that are supposed to show that an intention to deceive is not necessary for lying concentrate on testing bald-faced lies. However, these studies deliver mostly mixed results.[17] Some studies suggest that lay people agree that some bald-faced lies do not involve deception (Arico and Fallis 2013;

[15] Another example would involve purposeful misrepresentation of the Bohr model, and thus saying, for instance, that electrons travel in elliptical orbits, instead of circular. This is also a clear case of a lie, which shows that John is performing his explanation in a lie-prone context. Although this example may be seen as a more natural case of a lie, we use the unchanged Bohr model in both versions of *Atom Story* to focus only on the difference in the speaker's intention.

[16] See e.g. Turri and Turri 2015; Wiegmann et al. 2016; Marsili 2016; Viebahn et al. 2021; Reins and Wiegmann 2021; Gaszczyk 2022.

[17] Simultaneously, bald-faced lies have received two alternative theoretical treatments. Firstly, some argue that they are not genuine assertions and thus cannot be lies (see e.g. Meibauer 2014a, c; Dynel 2015; Leland 2015; Keiser 2016; Maitra 2018). Secondly, some argue that they are lies, granting a proper extension of an intention to deceive (see e.g. Lackey 2013; Meibauer 2014b; Rudnicki and Odrowąż-Sypniewska 2023).

Krstić and Wiegmann 2022; cf. Coleman and Kay 1981; Taylor et al. 2003). Other studies, however, deny that (Meibauer 2016; Rutschmann and Wiegmann 2017).[18] The question regarding an intention to deceive is crucial for our study which presents a pair of cases which differ only in the speaker's intention.

The empirical work on the notions of explanation and understanding corroborates their close relationship. Recent studies show that scientists and lay people alike consider a speech act to be an explanation when it provides an understanding to the audience (Waskan et al. 2014; Wilkenfeld and Lombrozo 2020). Wilkenfeld and Lombrozo (2020, 2590) argue that "… there is strong reason to believe not only that explanations are judged by the extent to which they produce some mental state (namely understanding), but that understanding is sufficient to play this mental state role." This is consistent with accounts of explanations that argue that the aim of explanation is to generate understanding in the audience (see e.g. Achinstein 1983; Wilkenfeld 2014; Gaszczyk 2023). Nevertheless, some questions remain unanswered. One of them concerns determining the minimal threshold for understanding. Some studies indicate that the threshold for understanding, in some respects, may be stronger than that for knowledge (Wilkenfeld et al. 2018). Yet, the exact way in which the notion of understanding should be elucidated remains open. In our study, we test stories in which it is explicitly stipulated that the given explanation is false. This is consistent with the assessment of such explanations in the literature (as discussed in Section 2.1). If the aim of explanation is to generate understanding, the question is whether the understanding in question can be non-factive. Our study can contribute towards answering this question.

2.4 The aim of the experimental studies

We conducted three experiments in which we presented to the participants two stories with false explanations. In each story, we have a teacher who tries to explain a complex phenomenon (for instance, the structure of an atom) to her students on a theory or a model that simplifies the phenomenon in question so that she knows that

[18] In all these studies people are asked whether someone is lying, not whether someone is a liar. We follow the same formula. Simultaneously, following the reviewer's observation, we acknowledge that asking the latter could influence ordinary speakers' judgements.

what she says is false. The vignettes have two conditions—positive intention and negative intention—that differ only in the speaker's intention. In the positive intention condition, the speaker makes a false explanation because she does not want to overwhelm her pupils or because she believes that it is pedagogically better to explain something that will be easier to understand. In the negative intention condition, the speaker gives a simplified (i.e., false) explanation because she wants to mislead the audience. The participants' task was to assess whether the speaker lied by making a false explanation.

Our prediction is that changing the speaker's intention influences the lie attribution. Namely, participants in the negative intention condition should more frequently claim that the speaker is lying than in the positive intention condition. The additional goal of our study is to test whether, depending on the speaker's intention, people would assess differently the degree to which the speaker explained a given phenomenon. Our prediction is that participants would be more inclined to agree that the speaker explained something in the positive intention condition than in the negative intention condition, even though in both conditions the explanation is the same.

3 Experiment 1

3.1 Method

3.1.1 Participants

One hundred and twenty U.S. and UK residents were tested, aged between 20 and 79 (M = 44.32; SD = 14.88); all with English as a native language; 54 male, 65 female; 1 prefer not to say). Because, according to our knowledge, this study is the first that tests experimentally lie attributions in false explanations, we decided to recruit a similar number of participants as in other studies in experimental philosophy using analogous tasks (e.g., Turri 2015a, b). In all experiments, participants were recruited and tested online using Qualtrics on Prolific Platform (Palan and Schitter 2018) and compensated £ 0.54 for about 4 min of their time. Repetitive participation was prevented. In all experiments, participation was voluntary, and participants were informed explicitly that they could withdraw their consent at any point during the study without any adverse consequences, and that their data would be anonymised.

All participants signed the informed consent form before the start of the experimental procedures. The experiments were reviewed and approved by the University of Groningen Research Ethics Review Committee (CETO ID 91217593).

3.1.2 Design, materials and procedure

Participants first read general instructions that familiarised them with the task. They were then randomly assigned to one of two conditions: positive intention (n = 60) versus negative intention (n = 60). In both conditions, participants were presented with two stories; one with a simplified explanation of the structure of an atom (*Atom Story*) and the other an explanation of a basic tenet of the theory of evolution (*Evolution Story*). Thus, experimental conditions (positive intentions vs. negative intentions) were manipulated between-subject and the topics of story were manipulated within-subject. *Atom Story* has already been introduced above, i.e., *Atom Story (positive intention)* and *Atom Story (negative intention)*. All other experimental materials are presented in Appendix. After reading each story, participants needed to assess the following statements: "X is lying," "X knows that his explanation is false," and "X explained Y" on a 6-point Likert scale (*1–strongly agree, 2–agree, 3–somewhat agree, 4–somewhat disagree, 5–disagree, 6–strongly disagree*).[19] There was no time limit for reading stories and assessing statements. Because each story contains information that the speaker was aware that their explanation is false, the second statement performs a function of manipulation check.

3.2 Results

First, we checked participants' lie attribution depending on the described speaker's intention (see Figure 1). These were analysed as a function of experimental condition (positive intention vs. negative intention) and story (*Atom Story* vs. *Evolution Story*),

---

[19] X represents the name of the speaker in each story and Y the explained phenomenon, e.g., the structure or an atom, or evolution. The statement "X knows that his explanation is false" was intended as a manipulation check, i.e., we intended to exclude the participants who disagreed with this statement from any further analyses. However, since the main effects were similar regardless of this exclusion, we decided to analyse the data based on the whole dataset. This not only increased the power of our analyses but also allowed us to keep a comparable number of participants per condition in each experiment.

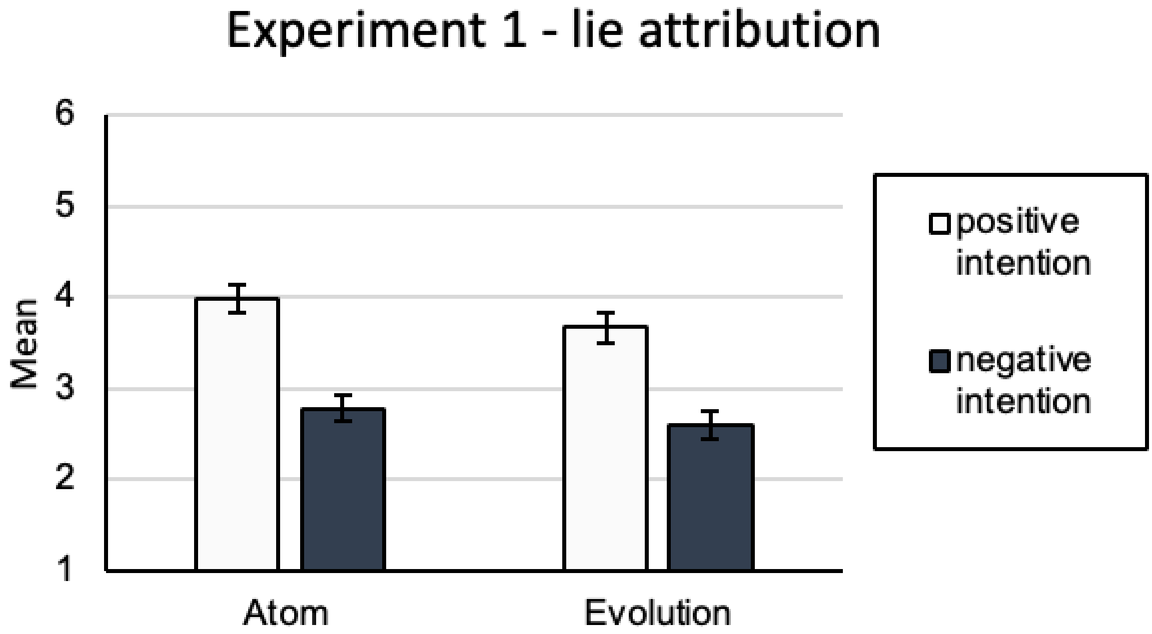
using a two-factor mixed ANOVA. The main effect of the experimental condition proved to be significant, $F(1, 118) = 39.55$, $MSE = 1.95$, $p < .001$, $\eta_p^2 = .251$. On average, participants were more inclined to claim that the speaker is lying in the negative intention condition ($M = 2.69$, $SD = 1.00$) than in the positive intention condition ($M = 3.83$, $SD = 0.97$).[20] A main effect of the story was also significant, $F(1, 118) = 4.38$, $MSE = 0.86$, $p = .038$, $\eta_p^2 = .036$. Participants were more inclined to agree that the speaker is lying in the case of *Evolution Story* ($M = 3.13$, $SD = 1.32$) than *Atom Story* ($M = 3.38$, $SD = 1.30$). The interaction between experimental conditions were insignificant, $F(1, 118) = 0.312$, $MSE = 0.86$, $p = .578$, $\eta_p^2 = .003$.

We also compared the number of participants who agree with the claim that the speaker is lying (i.e., who chose responses from *1–strongly agree* to *3–somewhat agree*) with the number of participants who disagree with this claim (i.e., who chose responses from *4–somewhat disagree* to *6–strongly disagree*). Significantly more participants claim that the speaker is lying in the negative intention condition than in the positive intention condition for both *Atom Story*: 70.0% vs 31.7%, $\chi^2(1) = 17.64$, $p < .001$, and *Evolution Story*: 78.3% vs 43.3%, $\chi^2(1) = 15.42$, $p < .001$.

[20] To emphasise, the numbers indicate how strongly participants disagree that the speaker lied. Thus, the smaller the number, the more they agree.

Figure 1

The degree to which participants attribute lying to the speaker.

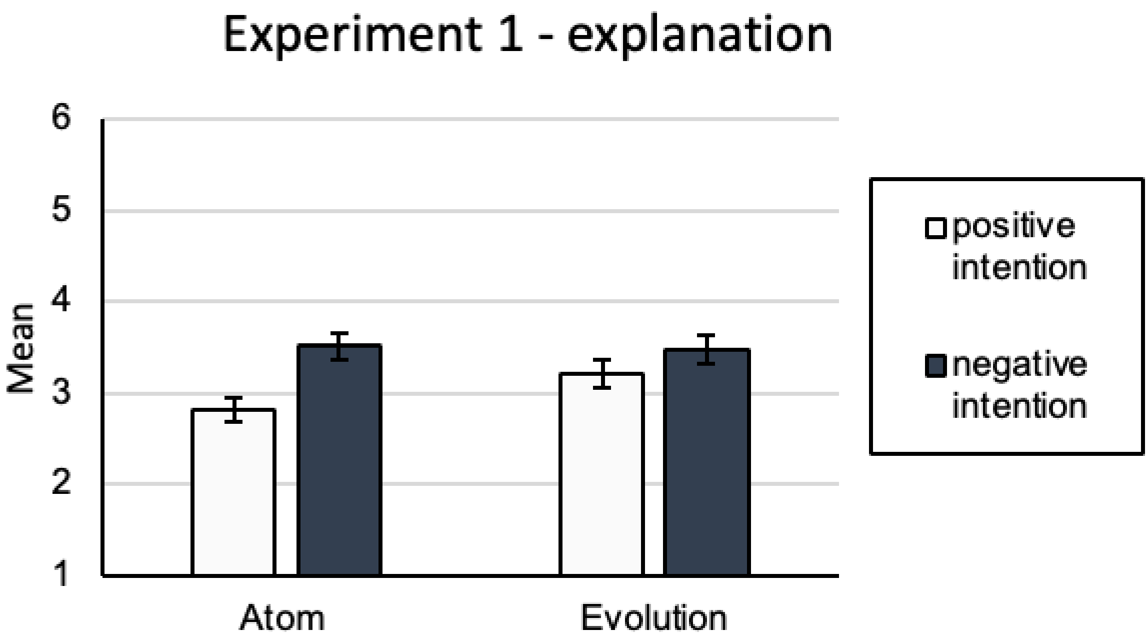


Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker lied.

Second, we investigated to what degree participants are inclined to say that the speaker explained the scientific phenomenon in question (see Figure 2). A 2 (condition) × 2 (story) ANOVA indicated a significant main effect of the condition, $F(1, 118) = 6.980$, $MSE = 2.01$, $p = .009$, $\eta_p^2 = .056$. Participants in the positive intention condition were more inclined to claim that the speaker explained given phenomena ($M = 3.02$, $SD = 1.00$) than in the negative intention condition ($M = 3.50$, $SD = 1.01$). A main effect of the story, $F(1, 118) = 3.96$, $MSE = 0.51$, $p = .049$, $\eta_p^2 = .032$, and an interaction, $F(1, 118) = 5.52$, $MSE = 0.51$, $p = .020$, $\eta_p^2 = .045$, also reached statistical significance. Pairwise comparisons with Bonferroni corrections showed that only in the case of the *Atom Story* there was a significant difference between positive and negative conditions, -0.70, $SE = 0.20$, $p < .001$. The assessment of the speaker's explanation in the *Evolution Story* was similar in positive and negative conditions, -0.27, $SE = 0.21$, $p = .211$.

Figure 2

The degree to which participants agree that the speaker explained a scientific phenomenon.



Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker explained the phenomenon.

When we grouped responses together, in *Atom Story*, more participants in the positive intention condition (80.0%) claimed that the speaker explained the phenomenon than in the negative intention condition (61.7%), $\chi^2(1) = 4.88$, $p = .027$. There was no such difference for *Evolution Story*, $\chi^2(1) = 0.33$, $p = .564$ (the positive intention condition: 68.3%; the negative intention condition: 63.3%).

Finally, we check the relationship between the lie attribution and the degree to which the participants agree that the speaker explained the phenomenon. The results were similar in both experimental conditions, so we grouped them together. The more participants claimed that the speaker was lying, the more they disagreed that they explained a given scientific phenomenon—moderate correlations: $r = -.541$, $p < .001$, for *Atom Story*, and $r = -.375$, $p < .001$, for *Evolution Story*.

3.3 Discussion

Does modification of the speaker's intention (positive/negative) change the lie attribution in the case of false explanations? The aim of Experiment 1 was to answer this question and, in doing so, test whether all such explanations should be classified as lies, as non-deceptionist definitions of lying predict. Our results go against this prediction. Participants distinguished between two conditions of the same stories—when the speaker's intention was deceitful, more participants classified the explanation as a lie than when the intention was positive. Moreover, participants attributed lying to a greater degree in the negative intention condition than in the positive intention condition. Still, a substantial number of participants agreed that the speakers were lying in the positive intention condition.

The second question concerned whether the stories are considered to be explanations. Here the results differ between the stories. In the case of *Atom Story*, participants distinguished between the positive and negative intention conditions—in the former, participants agreed to a greater extent that John explained the structure of an atom than in the latter. This shows that the change in intention influences not only the attribution of lying but also of explanation. The results for *Evolution Story* are less clear. In both conditions, participants similarly assessed the degree to which the speaker explained the phenomenon of evolution. Many factors may contribute to this result, such as participants' background knowledge or strong intuitions about the theory of evolution. It may also be important that participants are more eager to claim that the speaker is lying in *Evolution Story* than in *Atom Story*. Importantly, in both conditions, a negative correlation was observed between the degree of perceived explanation and lie attribution.

4 Experiment 2

The first aim of Experiment 2 was to test whether we would be able to replicate the observation regarding the lie attribution on new stories. Because the effect of the speaker's intention on lie attribution was not, according to our knowledge, tested experimentally before, it is of utmost importance to check whether the obtained pattern of results is not material driven. The second aim of Experiment 2 was to test whether we could receive more unequivocal results about the perceived degree of attributed explanation by the speaker. Finally, since one-third of the participants did not pass the manipulation check in the positive intention condition in Experiment 1, we wanted to test the overall robustness of the results.

4.1 Method

4.1.1 Participants

One hundred and twenty U.S. and UK residents aged between 22 and 77 years (*M* = 41.02; *SD* = 15.25) were tested; all with English as a native language; 89 female, 30 male). In Experiment 1, the main effect of the condition (positive intention vs. negative intention) on lie attribution had a very large effect size equal $\eta_p^2 = .251$. A total sample size equal to 14 participants would be sufficient to obtain a comparable effect with a statistical power of .95 and α level at .05. However, in Experiments 2 and 3, we decided to test about 60 participants per condition because we used new (Experiment 2) or modified versions (Experiment 3) of the stories and due to inevitably of more potential confounding variables during online testing.

4.1.2 Design, materials and procedure

The procedures were exactly the same as in Experiment 1. The difference was in the tested stories, i.e., *Solar System Story* and *Laws of Motion Story*, which can be found in the Appendix. To give the gist of the stories, in the former, the speaker explains the heliocentric model, by saying that the planets in our solar system move around the Sun in circular, instead of elliptical, orbits, even though he knows that this is a false depiction of our solar system. In the latter story, the speaker explains the laws of motion on the basis of Newton's laws, knowing that they are false and that we have much more accurate theories at our disposal.[21]

4.2 Results

We excluded one participant from the following analyses because their experimental data did not save due to a technical problem. We conducted 2 (condition) x 2 (story) mixed ANOVA to check participants' lie attribution (see Figure 3). There was a

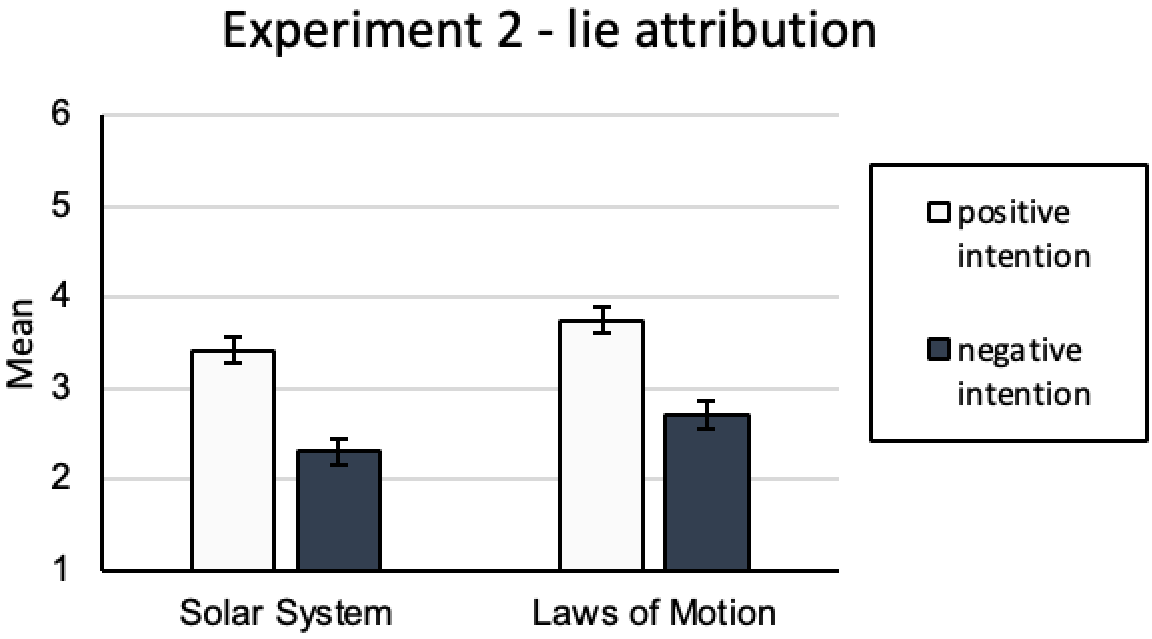
[21] While in the stories in Experiment 1, it is stated once that the explanation is false (both in the negative and in the positive intention conditions), in the negative intention conditions in Experiment 2, this fact is repeated. Nevertheless, we observed no difference in results between both experiments.

significant effect of condition, $F(1, 117) = 38.36$, $MSE = 1.79$, $p < .001$, $\eta_p^2 = .247$. Participants were more inclined to claim that the speaker is lying in the negative intention condition ($M = 2.51$, $SD = 1.01$) than in the positive intention condition ($M = 3.58$, $SD = 0.88$). The main effect of the story was also significant, $F(1, 117) = 11.52$, $MSE = 0.71$, $p < .001$, $\eta_p^2 = .090$. Participants were more eager to agree that the speaker is lying in *Solar System Story* ($M = 2.87$, $SD = 1.21$) than *Laws of Motion Story* ($M = 3.24$, $SD = 1.26$). The interaction was not significant, $F(1, 117) = 0.11$, $MSE = 0.71$, $p = .737$, $\eta_p^2 = .001$.

A binary comparison shows that significantly more participants agree that the speaker is lying in the negative intention condition than in the positive intention condition for both *Solar System Story*: 89.8% vs 56.7%, $\chi^2(1) = 16.64$, $p < .001$, and *Laws of Motion Story*: 78.0% vs 36.7%, $\chi^2(1) = 20.72$, $p < .001$.

Figure 3

The degree to which participants attribute lying to the speaker.

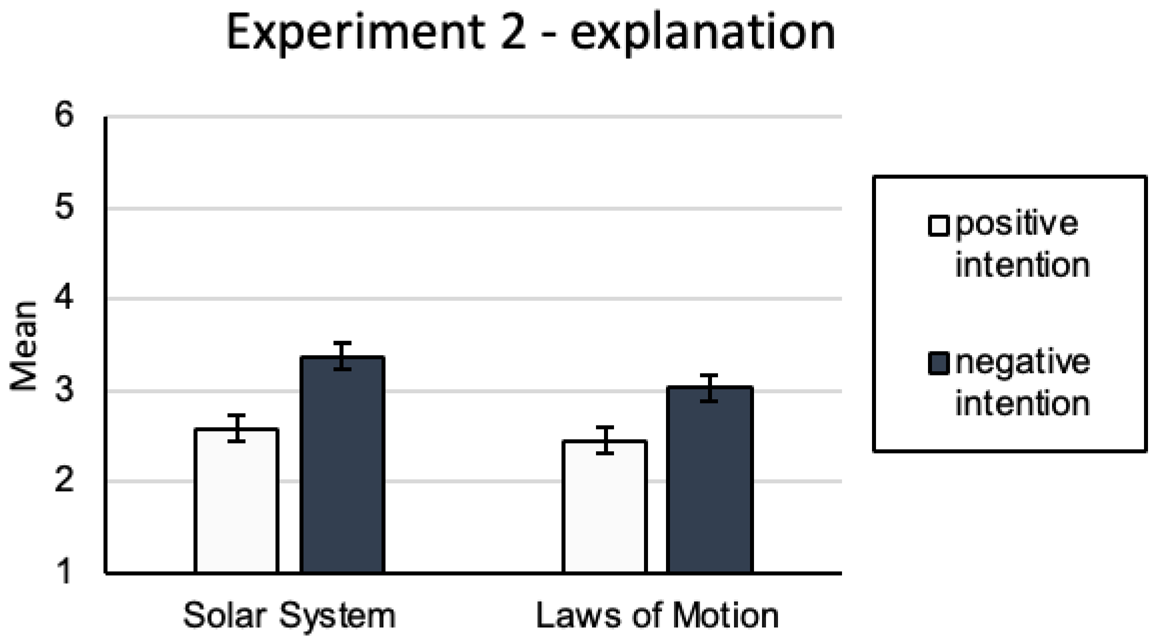


Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker lied.

Next, we performed a 2 (condition) × 2 (story) ANOVA on participants' assessment of explanations provided by the speakers (see Figure 4). The ANOVA showed a significant main effect of the condition, $F(1, 117) = 15.93$, $MSE = 1.76$, $p < .001$, $\eta_p^2 = .120$ and the story, $F(1, 117) = 4.91$, $MSE = 0.68$, $p = .029$, $\eta_p^2 = .040$. In the positive intention condition, participants were more inclined to claim that the speaker explained the given phenomenon ($M = 2.53$, $SD = 0.84$) than in the negative intention condition ($M = 3.27$, $SD = 1.02$). Moreover, reading the *Laws of Motion Story* ($M = 2.74$, $SD = 1.13$), participants claimed to a greater degree that the speaker explained the scientific phenomenon than when reading *Solar System Story* ($M = 2.97$, $SD = 1.18$). The interaction was not significant, $F(1, 117) = 0.93$, $MSE = 0.68$, $p < .337$, $\eta_p^2 = .008$. These results are consistent also with χ^2 analysis when we combined *agree/disagree* responses. More participants agree that the speaker explained the scientific phenomenon in the positive intention condition than in the negative intention condition, for both *Solar System Story*: 90.0% vs 66.1%, $\chi^2(1) = 9.95$, $p = .002$, and *Laws of Motion Story*: 86.7% vs 74.6%, $\chi^2(1) = 2.79$, $p = .095$. However, in the latter case, the difference was only numerical.

Figure 4

The degree to which participants agree that the speaker explained a scientific phenomenon.



Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker explained the phenomenon.

Finally, as in Experiment 1, we checked the relationship between lie attribution and explanation attribution. Again, the more participants claimed that the speaker was lying, the more they disagreed that the speaker explained the scientific phenomenon—moderate correlations: $r = -.317$, $p < .001$, for *Solar System Story*, and $r = -.474$, p < .001, for *Laws of Motion Story*.

4.3 Discussion

The results replicated the main findings from Experiment 1. When the speaker intended to mislead the audience, participants to a greater degree classified the explanation as a lie than when the speaker's intention was positive. This difference between the conditions is hard to explain by non-deceptionists. Simultaneously, again, a substantial number of participants, especially in *Solar System Story*, agreed that the speakers were lying in the positive intention condition. When it comes to the judgments of whether the stories are considered to be explanations, this time in both stories, participants distinguished between the positive and negative intention conditions—in the positive intention conditions, participants were more inclined to agree that the speakers explained the phenomena than in the negative intention conditions. Again, there was a negative correlation between the perceived degree of explanation and lying.

It is worth noting that participants were more willing to agree that the speaker is lying and disagree that the speaker explained the given phenomenon while reading *Solar System Story* than *Law of Motion Story.* These differences may stem from story-specific qualities. For example, participants may have assumed that it would be relatively easy to provide the correct explanation (use elliptical orbits instead of circular ones) in the case of *Solar System Story*.

5 Experiment 3

The aim of Experiment 3 was to extend the observations from Experiments 1 and 2 in two important aspects. Firstly, we added a new experimental condition, in which the speaker's intention was unspecified, to test whether there was any difference between the negative intention and the lack of intention in educational contexts. It is possible that in such contexts, lay people assume the positive intention of the teacher. If so, we would observe no difference between the positive intention and the lack of intention conditions. At the same time, participants would be more inclined to claim that the speaker is lying in the negative intention than in the lack of intention condition.

The second objective of Experiment 3 was to give participants an option to withhold the decision of whether they agree/disagree with the assessed statements. To this end, we added the option "neither agree nor disagree" to the Likert scale. We wanted to preclude the possibility that the pattern of results in the previous experiments emerged mostly because the participants were forced to agree or disagree. Thus, we wanted to see whether we were able to replicate the previous results on the modified scale.[22]

5.1 Method

5.1.1 Participants

In total, 179 participants aged between 20 and 79 (M = 40.70; SD = 13.13) were tested; (125 females, 3 non-binary or gender diverse, 50 males, 1 prefer not to say).

5.1.2 Design, materials and procedure

In Experiment 3, the design and procedure were similar to those in Experiments 1 and 2. However, compared to the procedure in the previous experiments, this time participants were randomly assigned to one of three conditions: positive intention (n = 59), negative intention (n = 61), or lack of intention (n = 58). In all conditions, participants were presented with two stories—one from each of the previous experiments: *Atom Story* and *Laws of Motion Story*.[23] Participants assess the same statements as in the previous experiments, but this time using a 7-point Likert scale (*1–strongly agree, 2–agree, 3–somewhat agree, 4-neither agree or disagree, 5–somewhat disagree, 6–disagree, 7–strongly disagree*).

5.2. Results

As in previous experiments, we checked participants' lie attribution (see Figure 5). These were analysed in the 3 (condition) × 2 (story) mixed ANOVA. Again, the ANOVA yielded a significant effect of condition, $F(2, 176) = 17.39$, $MSE = 4.03$, $p < .001$, $\eta_p^2 = .165$. Pairwise comparisons with Bonferroni corrections showed that, as in previous experiments, participants were more inclined to claim that the speaker is

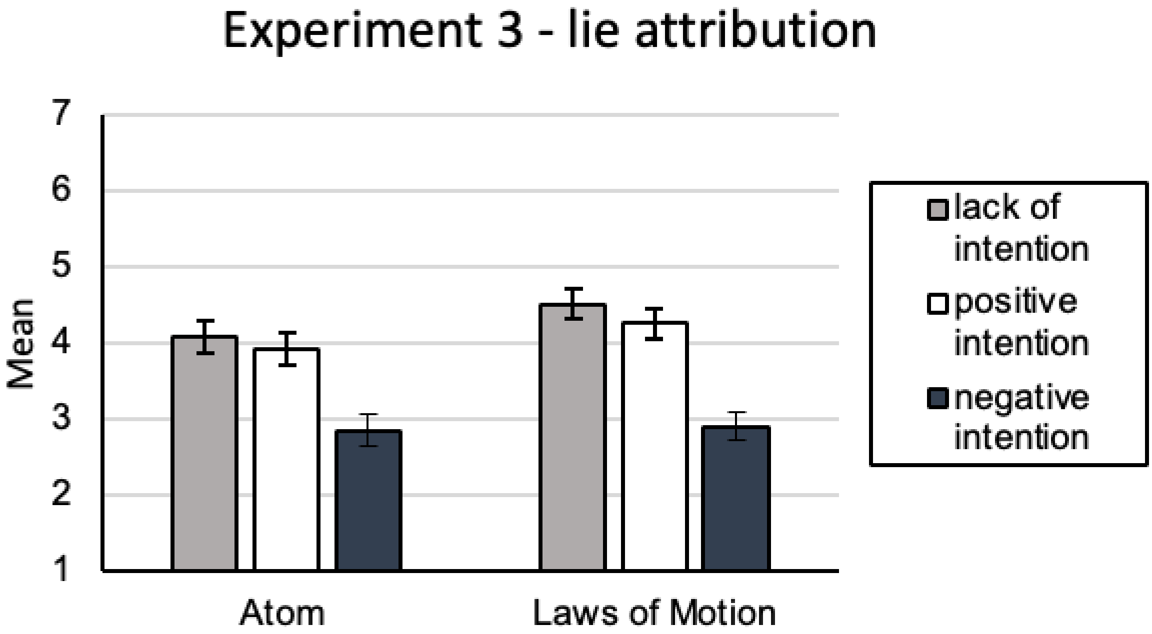
[22] We thank the anonymous reviewers for these research suggestions.

[23] One modification in *Atom Story* is that we are now saying that "John knows that this is a false depiction of an atom." We omitted "a crude simplification" from this statement. Even though saying that an explanation is both a simplification and a falsity is often put together by non-factivists (e.g. Elgin 2007), we decided to disambiguate this statement. We stipulate that the presented explanations are false because this is how they are explicated in the debate—their falsity is agreed upon and uncontroversial (see references in footnote 1).

lying in the negative intention condition (M = 2.88, SD = 1.29) than in the positive intention condition (M = 4.08, SD = 1.54), -1.21, $p < .001$. Similarly, more participants were willing to attribute lying to the speaker in the negative than neutral condition (M = 4.29, SD = 1.42), -1.41, $p < .001$. However, the positive and neutral conditions did not differ in that regard, -0.20, $p = 1.00$. As in Experiment 2, the main effect of the story was also significant, $F(1, 176) = 8.63$, $MSE = 0.79$, $p = .004$, $\eta_p^2 = .047$. This time, participants were more eager to agree that the speaker is lying in *Atom Story* (M = 3.60, SD = 1.70) than in *Laws of Motion Story* (M = 3.88, SD = 1.64). The interaction was not significant, $F(2, 176) = 1.57$, $MSE = 0.79$, $p = .211$, $\eta_p^2 = .08$.

Figure 5

The degree to which participants attribute lying to the speaker.

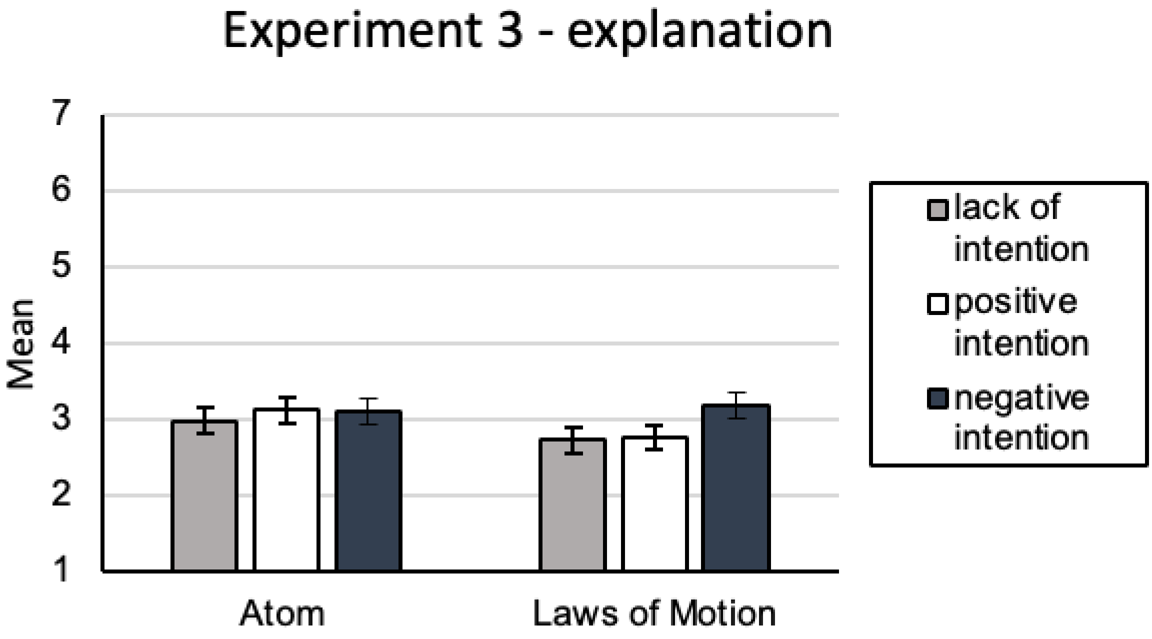


Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker lied.

Participants' agreement that the speaker explained the scientific phenomena indicated a marginally significant main effect of the story, $F(1, 176) = 3.26$, $MSE = 0.85$, $p = .073$, $\eta_p^2 = .018$; participants were slightly more inclined to claim that the speaker explained the scientific phenomena in the case of *Laws of Motion Story* ($M = 2.89$, $SD = 1.32$) than in *Atom Story* ($M = 3.07$, $SD = 1.28$). The main effect of the condition, $F(2, 176) = 1.01$, $MSE = 2.51$, $p = .365$, $\eta_p^2 = .011$, and the interaction were insignificant, $F(2, 176) = 1.86$, $MSE = 0.85$, $p = .159$, $\eta_p^2 = .021$.

Figure 6

The degree to which participants agree that the speaker explained a scientific phenomenon.



Note. Error bars represent standard errors. The numbers indicate how strongly participants disagree that the speaker explained the phenomenon.

As in previous experiments we observed significant and negative correlations between lie attribution and explanation attribution: r = -.381, p < .001, for *Atom Story*, and r = -.437, p < .001, for *Laws of Motion Story*.

5.3 Discussion

In Experiment 3, participants claimed to a greater degree that the speaker is lying in the negative intention condition than in both the positive intention condition and the lack of intention condition. Moreover, the last two conditions did not differ in terms of lying assessment. Thus, we not only replicated the results observed in Experiments 2 and 3 but also extended these observations to the condition where the intention was not provided. Thus, it seems that providing information about a positive intention in an educational domain is not necessary because teachers' trustworthiness is assessed similarly regardless of knowing their positive intentions or not. This changes when the speaker's intention is deceptive.

In all experimental conditions, participants' assessment of whether the speaker explained the scientific phenomena was similar. This result is in contrast with observations from Experiment 2 (and partially with Experiment 1) where the participants claimed that the speaker explained a scientific phenomenon to a greater degree in the positive intention condition than in the negative intention condition. When we combined responses "strongly agree," "agree" and "slightly agree" together, about 80% of participants in each condition chose one of these options.[24] In the previous experiments, it was approximately 65%. Thus, the observed lack of difference between the conditions may stem from the fact that the participants generally claimed that the speaker explained the scientific phenomenon.

6 General discussion

The aim of the experiments was to test the judgements of ordinary speakers regarding lie attributions in false explanations. Our prediction was that changing the speaker's intention would influence the lie attribution in such explanations. Simultaneously, the prediction of non-deceptionists, the dominant way of defining lying, is that the speaker lies by making such explanations, independently of her intention. The results challenged this prediction. The experiments demonstrated that the speaker's intention matters to lay people. Across four different scenarios, participants agreed to a higher degree that the speaker lies in the negative intention condition than in the positive intention condition.

Our prediction was also that participants would be more inclined to agree that the speaker explained something in the positive intention condition than in the negative intention condition, even though in both conditions the explanation was the same. We suspected that the speaker's intention matters not only for lie attribution but influences also other factors. In Experiments 1 and 2, in all cases except *Evolution Story*, participants were more inclined to attribute explanation in the positive intention condition than in the negative intention condition. In the theoretical part,

[24] For *Atom Story* it was 74.6% in the positive intention condition, 70.5% in the negative intention condition, and 84.7% in the lack of intention condition. For *Laws of Motion Story*, it was 83.1% in the positive intention condition, 72.1% in the negative intention condition, and 84.7% in the lack of intention condition.

we proposed that such false explanations should be considered genuine explanations and that they involve non-factive understanding. Still, some participants assessed that the speaker did not perform an explanation (although it was only around 20% in Experiment 3). Previous studies (e.g. Wilkenfeld et al. 2018) indicated that lay people's notions of explanation and understanding have minimal thresholds, however, so far, these thresholds have not been specified. If the speakers do not explain something, some may wonder, what kind of speech acts they are performing. One possibility is that lay people's notion of explanation is more exclusive, and thus false explanations belong to a different, broader category, like teaching something. However, in a study that directly addresses the question of what kind of speech acts false explanations are, it may turn out that lay people do classify them as genuine explanations.[25] Nevertheless, our main question did not consider the nature of explanation and understanding, rather we focused only on explanations in relation to the lie attribution. Thus, we leave this topic for further research in the future.

Non-deceptionists fail to distinguish between two conditions (positive intention vs. negative intention) of the stories and thus they incorrectly classify both conditions as lies (the same concerns the lack of intention condition). Our stance is the following: non-deceptionists are on the right track in defining lying as insincerely asserting, however, at least in some cases, there is a need for a further criterion in the definition of lying.

One way out is reintroducing, appropriately modified, an intention to deceive in a definition of lying.[26] Consider, for instance, Lackey's (2013) deceptionist

[25] This is the case with the experimental investigation of whether lies must be false. Turri and Turri (2015) present empirical evidence for this claim, thus excluding the possibility of true lies. However, recent experimental data, failing to replicate their study, give evidence to the contrary, see e.g. Wiegmann et al. 2016.

[26] We do not claim that any version of a deceptionist definition of lying distinguishes between these conditions. Consider, for instance, that if a deceptionist would say that lying is saying something one believes to be false with an attempt to make the audience believe something that one does not believe, then John is lying both in positive and negative conditions. Crucially, there are more ways in which an intention to deceive can be explicated, see e.g. Mahon 2016; Krstić 2023. In the main text, we give an example of a definition that can correctly handle the cases. Since the difference in the presented stories lies in the speakers' intentions, we see deceptionism as the most natural option for explaining the difference between the conditions.

definition of lying. She argues that lying involves intending to be deceptive, which she understands as aiming to conceal information from the audience.[27] John in *Atom Story (positive intention)* has no intention to be deceptive in this sense, even though he says something he believes to be false; in contrast, in *Atom Story (negative intention)* he intends to deceive his audience. Thus, it seems that Lackey's definition delivers a desirable result. Accepting a deceptionist definition—due to bald-faced lies—is very controversial, however.[28] Another option is to propose a definition of lying that is not formulated in a deceptionist way but would simultaneously allow us to make correct predictions. Peet's (2021) definition seems to do the job here and can be seen as a compromise between deceptionists and non-deceptionists. For Peet, lying involves displaying poor testimonial worth in one's insincere assertion, where 'testimonial worth' is understood as "… a reflection of the quality of character a speaker displays in asserting" (2021, 2392). A poor testimonial worth does not necessarily mean that one's assertion is false, rather it involves the assessment of the general character of the speaker as a testifier. Looking at John, in *Atom Story (positive intention)* he displays good testimonial worth since he makes a false explanation to not confuse and overwhelm his audience. By contrast, in *Atom Story (negative intention)* he gives the same explanation to deceive his audience, and thus displays poor testimonial worth. These two definitions are only examples of the recent accounts of lying that distinguish between both conditions. Our aim in this paper was not to provide a definition of lying but to present new data that need to be taken under consideration. More work must be done to propose a definition of lying that will be able to correctly capture these and all other cases.

Nevertheless, one may think of construing a non-deceptionist definition of lying that could, at least for some cases, deliver correct predictions.[29] However, such a definition cannot merely say that lies are insincere assertions. Rather, it should employ some additional resources. Here are two possibilities. Firstly, a

[27] There is a growing literature on the difference between concealing information, which counts as deceptive, and, e.g., withholding information, which is not necessarily deceptive; see e.g. Fallis 2020; Dynel 2020.

[28] Lackey's (2013) proposal received some pushback, see e.g. Fallis 2015; Stokke 2018; but see also footnote 17 for literature on alternative treatments of bald-faced lies.

[29] We thank an anonymous reviewer for asking us to elaborate on this point.

non-deceptionist could try to explicate the analysed explanations as cases of loose talk, or employ a scalar notion of sincerity.[30] Thus, for instance, one could argue that the explanation in the positive intention condition is sincere because it is true enough or it meets a certain threshold on a sincerity scale, while the explanation in the negative intention condition does not satisfy these standards and so it is insincere. Secondly, a non-deceptionist could propose that there are two standards of sincerity for explanations—one should not only believe (or express one's belief) that *p* but also understand *p*.[31] Here, some could argue that while the speaker in the positive intention condition is violating the belief standard but complies with the understanding standard, the speaker in the negative intention condition is violating both standards of sincerity. However, these two proposals need to be fully developed to see whether they can satisfactorily address the tested cases.

As a final note, we want to point to two potential directions for future research. Both concern the ideas mentioned above. Firstly, we only tested vignettes that explicitly stated that the explanations were false. However, it is worth checking if the participants' judgments change when alternative formulations are used, such as saying that the explanation is true enough, is a simplification, or is a case of loose talk. Secondly, we focused on the belief standard of sincerity, however, there may also be an additional sincerity standard, one based on understanding. It would be worth investigating how these two relate to each other, whether the speaker can violate one without the other, or whether such violations result in insincerity or lying. We hope that our study can be seen as a stepping-stone in researching the relationship between the concepts of lying and sincerity, on the one hand, and explanation and understanding, on the other.

[30] For views that treat sincerity and lying as gradable notions, see e.g. Carson 2006; Marsili 2014; Egré and Icard 2018; there is a rich literature on the notion of loose talk that originates in the work of Grice, for most recent examples, see e.g. Hoek 2018; Carter 2019; Moss 2019. Employing the ideas of scalarity of sincerity and loose talk could be particularly useful to analyse undiscussed cases, like cooperative attempts of explanations by someone who understands something insufficiently to explain it, or explanations in which false statements are on the periphery.

[31] It is worth noting that it has been argued that we can lie by expressing attitudes beyond beliefs. In his experimental study, Marsili (2016) claims that one can be insincere by promising that *p* if one believes that not *p* or intends not to *p* (or by both). Marsili's study aims to establish that we can lie with other attitudes than beliefs. In contrast, in our study, we do not address the question of the speaker's attitudes, particularly whether the speaker who explains something is insincere due to the lack of understanding. What we focus on is whether the speaker lies when explaining something false. See also footnote 10.

Acknowledgements

Many thanks to two anonymous referees for this journal, J. P. Grodniewicz, Emar Maier, and Elizabeth Maylor for very helpful comments and discussion, and to Marcin Danielewski for programming the experimental procedure for Experiment 3.

Declarations

Conflicts of interest/Competing interests: Not applicable

References

Achinstein, P. (1983). *The Nature of Explanation*. Oxford University Press.

Arico, A. J., & Fallis, D. (2013). Lies, Damned Lies, and Statistics: An Empirical Investigation of the Concept of Lying. *Philosophical Psychology*, *26*(6), 790–816. https://doi.org/10.1080/09515089.2012.725977

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.

Baumberger, C., Beisbart, C., & Brun, G. (2017). What is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (pp. 1–34). Routledge. https://doi.org/10.4324/9781315686110

Baumberger, C., & Brun, G. (2017). Dimensions of Objectual Understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (pp. 165–189). Routledge. https://doi.org/10.4324/9781315686110

Carson, T. L. (2006). The Definition of Lying. *Noûs*, *40*(2), 284–306. https://doi.org/10.1111/j.0029-4624.2006.00610.x

Carson, T. L. (2010). *Lying and Deception: Theory and Practice*. Oxford University Press.

Carter, S. (2019). The Dynamics of Loose Talk. *Noûs*, *55*(1), 171–198. https://doi.org/10.1111/nous.12306

Coleman, L., & Kay, P. (1981). Prototype Semantics: The English Word Lie. *Language*, *57*(1), 26–44. https://doi.org/10.1353/lan.1981.0002

De Regt, H. W. (2009). Understanding and Scientific Explanation. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives* (pp. 21–42). Pittsburgh University Press.

De Regt, H. W. (2017). *Understanding Scientific Understanding*. Oxford University Press.

De Regt, H. W., & Dieks, D. (2005). A Contextual Approach to Scientific Understanding. *Synthese*, *144*(1), 137–170. https://doi.org/10.1007/s11229-005-5000-4

De Regt, H. W., & Gijsbers, V. (2017). How False Theories Can Yield Genuine Understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (pp. 50–75). Routledge. https://doi.org/10.4324/9781315686110

Dynel, M. (2015). Intention to Deceive, Bald-faced Lies, and Deceptive Implicature: Insights into Lying at the Semantics-Pragmatics Interface. *Intercultural Pragmatics*, *12*(3). https://doi.org/10.1515/ip-2015-0016

Dynel, M. (2020). To Say the Least: Where Deceptively Withholding Information Ends and Lying Begins. *Topics in Cognitive Science*, *12*(2), 555–582. https://doi.org/10.1111/tops.12379

Egré, P., & Icard, B. (2018). Lying and Vagueness. In J. Meibauer (Ed.), *The Oxford Handbook of Lying* (pp. 353–369). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198736578.013.27

Elgin, C. Z. (2004). True Enough. *Philosophical Issues*, *14*(1), 113–131. https://doi.org/10.1111/j.1533-6077.2004.00023.x

Elgin, C. Z. (2007). Understanding and the Facts. *Philosophical Studies*, *132*(1), 33–42. https://doi.org/10.1007/s11098-006-9054-z

Elgin, C. Z. (2009). Is Understanding Factive? In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic Value* (pp. 322–330). Oxford: Oxford University Press.

Elgin, C. Z. (2017). *True Enough*. MIT Press.

Fallis, D. (2009). What Is Lying? *The Journal of Philosophy*, *106*(1), 29–56. https://doi.org/10.5840/jphil200910612

Fallis, D. (2012). Lying as a Violation of Grice's First Maxim of Quality. *Dialectica*, *66*(4), 563–581. https://doi.org/10.1111/1746-8361.12007

Fallis, D. (2013). Davidson was Almost Right about Lying. *Australasian Journal of Philosophy*, *91*(2), 337–353. https://doi.org/10.1080/00048402.2012.688980

Fallis, D. (2015). What Is Disinformation? *Library Trends*, *63*(3), 401–426. https://doi.org/10.1353/lib.2015.0014

Fallis, D. (2020). Shedding Light on Keeping People in the Dark. *Topics in Cognitive Science*, *12*(2), 535–554. https://doi.org/10.1111/tops.12361

García-Carpintero, M. (2021). Lying vs. Misleading: The Adverbial Account. *Intercultural Pragmatics*, *18*(3), 391–413. https://doi.org/10.1515/ip-2021-2011

Gaszczyk, G. (2019). Are Selfless Assertions Hedged? *Rivista Italiana Di Filosofia Del Linguaggio*, *13*(1), 47–54. https://doi.org/10.4396/09201909

Gaszczyk, G. (2022). Lying with Uninformative Speech Acts. *Canadian Journal of Philosophy*, *52*(7), 746–760. https://doi.org/10.1017/can.2023.12

Gaszczyk, G. (2023). Helping Others to Understand: A Normative Account of the Speech Act of Explanation. *Topoi*, *42*, 385–396. https://doi.org/10.1007/s11245-022-09878-y

Goldberg, S. (2015). *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford University Press.

Greco, J. (2014). Episteme: Knowledge and Understanding. In K. Timpe & C. A. Boyd (Eds.), *Virtues and Their Vices* (pp. 285–301). Oxford University Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.

Grimm, S. R. (2006). Is Understanding a Species of Knowledge? *The British Journal for the Philosophy of Science*, *57*(3), 515–535.

Grimm, S. R. (2010). The Goal of Explanation. *Studies in History and Philosophy of Science Part A*, *41*(4), 337–344. https://doi.org/10.1016/j.shpsa.2010.10.006

Grimm, S. R. (2016). How Understanding People Differs from Understanding the Natural World. *Philosophical Issues*, *26*(1), 209–225. https://doi.org/10.1111/phis.12068

Hills, A. (2016). Understanding Why: Understanding Why. *Noûs*, *50*(4), 661–688. https://doi.org/10.1111/nous.12092

Hoek, D. (2018). Conversational Exculpature. *The Philosophical Review*, *127*(2), 151–196. https://doi.org/10.1215/00318108-4326594

Isenberg, A. (1964). Deontology and the Ethics of Lying. *Philosophy and Phenomenological Research*, *24*(4), 463–480. https://doi.org/10.2307/2104756

Jäger, C. (2016). Epistemic Authority, Preemptive Reasons, and Understanding. *Episteme*, *13*(2), 167–185. https://doi.org/10.1017/epi.2015.38

Keiser, J. (2016). Bald-Faced Lies: How to Make a Move in a Language Game Without Making a Move in a Conversation. *Philosophical Studies*, *173*(2), 461–477. https://doi.org/10.1007/s11098-015-0502-5

Kelp, C. (2015). Understanding Phenomena. *Synthese*, *192*(12), 3799–3816. https://doi.org/10.1007/s11229-014-0616-x

Kelp, C., & Simion, M. (2021). *Sharing Knowledge: A Functionalist Account of Assertion*. Cambridge University Press. https://doi.org/10.1017/9781009036818

Khalifa, K. (2012). Inaugurating Understanding or Repackaging Explanation? *Philosophy of Science*, *79*(1), 15–37. https://doi.org/10.1086/663235

Khalifa, K. (2013). Understanding, Grasping, and Luck. *Episteme*, *10*(1), 1–17. https://doi.org/10.1017/epi.2013.6

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge University Press.

Krstić, V. (2019). Can You Lie Without Intending to Deceive? *Pacific Philosophical Quarterly*, *100*(2), 642–660. https://doi.org/10.1111/papq.12241

Krstić, V. (2023). Lying: Revisiting the 'intending to deceive' condition. *Analysis*, anac099. https://doi.org/10.1093/analys/anac099

Krstić, V., & Wiegmann, A. (2022). Bald-Faced Lies, Blushing, and Noses that Grow: An Experimental Analysis. *Erkenntnis*. https://doi.org/10.1007/s10670-022-00541-x

Kvanvig, J. L. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9780511498909

Lackey, J. (2013). Lies and Deception: An Unhappy Divorce. *Analysis*, *73*(2), 236–248. https://doi.org/10.1093/analys/ant006

Lipton, P. (2004). *Inference to the Best Explanation* (2nd ed.). Routledge. https://www.routledge.com/Inference-to-the-Best-Explanation/Lipton/p/book/9780415242035

Mahon, J. E. (2008). Two Definitions of Lying. *International Journal of Applied Philosophy*, *22*(2), 211–230. https://doi.org/10.5840/ijap200822216

Mahon, J. E. (2016). The Definition of Lying and Deception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/lying-definition/

Maitra, I. (2018). Lying, Acting, and Asserting. In Michaelson, Eliot & Stokke, Andreas (Eds.), *Lying: Language, Knowledge, Ethics, and Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780198743965.003.0004

Marsili, N. (2014). Lying as a Scalar Phenomenon: Insincerity along the Certainty-Uncertainty Continuum. In S. Cantarini, W. Abraham, & E. Leiss (Eds.), *Certainty-uncertainty – and the Attitudinal Space in Between* (Vol. 165, pp. 153–173). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.165.09mar

Marsili, N. (2016). Lying by Promising. *International Review of Pragmatics*, *8*(2), 271–313. https://doi.org/10.1163/18773109-00802005

Marsili, N. (2020). Lying, Speech Acts, and Commitment. *Synthese*, *199*(1–2), 3245–3269. https://doi.org/10.1007/s11229-020-02933-4

Marsili, N., & Löhr, G. (2022). Saying, Commitment, and the Lying-Misleading Distinction. *The Journal of Philosophy*, *119*(12), 687–698. https://doi.org/10.5840/jphil20221191243

McKagan, S. B., Perkins, K. K., & Wieman, C. E. (2008). Why We Should Teach the Bohr Model and How to Teach it Effectively. *Physical Review Special Topics - Physics Education Research*, *4*(1), 010103. https://doi.org/10.1103/PhysRevSTPER.4.010103

McKinnon, R. (2015). *The Norms of Assertion: Truth, Lies, and Warrant*. Palgrave Macmillan.

Meibauer, J. (2014a). A Truth That's Told with Bad Intent: Lying and Implicit Content. *Belgian Journal of Linguistics*, *28*(1), 97–118. https://doi.org/10.1075/bjl.28.05mei

Meibauer, J. (2014b). Bald-Faced Lies as Acts of Verbal Aggression. *Journal of Language Aggression and Conflict*, *2*(1), 127–150. https://doi.org/10.1075/jlac.2.1.05mei

Meibauer, J. (2014c). *Lying at the Semantics-Pragmatics Interface*. De Gruyter Mouton. 10.1515/9781614510840

Meibauer, J. (2016). Understanding Bald-Faced Lies. *International Review of Pragmatics*, *8*(2), 247–270. https://doi.org/10.1163/18773109-00802004

Milić, I. (2017). Against Selfless Assertions. *Philosophical Studies*, *174*(9), 2277–2295. https://doi.org/10.1007/s11098-016-0798-9

Moss, S. (2019). Full Belief and Loose Speech. *The Philosophical Review*, *128*(3), 255–291. https://doi.org/10.1215/00318108-7537270

Palan, S., & Schitter, C. (2018). Prolific.ac—A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peet, A. (2021). Testimonial Worth. *Synthese*, *198*(3), 2391–2411. https://doi.org/10.1007/s11229-019-02219-4

Pepp, J. (2022). What Is the Commitment in Lying. *The Journal of Philosophy*, *119*(12), 673–686. https://doi.org/10.5840/jphil20221191242

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/I/bo27128726.html

Pratchett, T., Stewart, I., & Cohen, J. (1999). *The Science of Discworld*. Ebury Press.

Primoratz, I. (1984). Lying and the 'Methods of Ethics'. *International Studies in Philosophy*, *16*(3), 35–57. https://doi.org/10.5840/intstudphil198416353

Reins, L. M., & Wiegmann, A. (2021). Is Lying Bound to Commitment? Empirically Investigating Deceptive Presuppositions, Implicatures, and Actions. *Cognitive Science*, *45*(2). https://doi.org/10.1111/cogs.12936

Rudnicki, J., & Odrowąż-Sypniewska, J. (2023). Don't be deceived: Bald-faced lies are deceitful assertions. *Synthese*, *201*(6), 192. https://doi.org/10.1007/s11229-023-04180-9

Rutschmann, R., & Wiegmann, A. (2017). No Need for an Intention to Deceive? Challenging the Traditional Definition of Lying. *Philosophical Psychology*, *30*(4), 438–457. https://doi.org/10.1080/09515089.2016.1277382

Saul, J. M. (2012). *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford University Press. 10.1093/acprof:oso/9780199603688.001.0001

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Sorensen, R. (2007). Bald-Faced Lies! Lying Without the Intent to Deceive. *Pacific Philosophical Quarterly*, *88*(2), 251–264. https://doi.org/10.1111/j.1468-0114.2007.00290.x

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics* (Vol. 9). Academic Press; (References are to the reprint in R. Stalnaker, Context and content. Oxford University Press. 1999. pp. 78–95.).

Stokke, A. (2014). Insincerity. *Noûs*, *48*(3), 496–520. https://doi.org/10.1111/nous.12001

Stokke, A. (2018). *Lying and Insincerity*. Oxford University Press. 10.1093/oso/9780198825968.001.0001

Strevens, M. (2013). No Understanding without Explanation. *Studies in History and Philosophy of Science Part A*, *44*(3), 510–515. https://doi.org/10.1016/j.shpsa.2012.12.005

Taylor, M., Lussier, G. L., & Maring, B. L. (2003). The Distinction Between Lying and Pretending. *Journal of Cognition and Development*, *4*(3), 299–323. https://doi.org/10.1207/S15327647JCD0403_04

Turri, A., & Turri, J. (2015). The Truth about Lying. *Cognition*, *138*, 161–168. https://doi.org/10.1016/j.cognition.2015.01.007

Turri, J. (2015a). Assertion and Assurance: Some Empirical Evidence. *Philosophy and Phenomenological Research*, *90*(1), 214–222. https://doi.org/10.1111/phpr.12160

Turri, J. (2015b). Selfless Assertions: Some Empirical Evidence. *Synthese*, *192*(4), 1221–1233. https://doi.org/10.1007/s11229-014-0621-0

Turri, J. (2015c). Understanding and the Norm of Explanation. *Philosophia*, *43*(4), 1171–1175. https://doi.org/10.1007/s11406-015-9655-x

Viebahn, E. (2020). Lying with Presuppositions. *Noûs*, *54*(3), 731–751. https://doi.org/10.1111/nous.12282

Viebahn, E. (2021). The Lying-Misleading Distinction: A Commitment-Based Approach. *The Journal of Philosophy*, *118*(6), 289–319. https://doi.org/10.5840/jphil2021118621

Viebahn, E., Wiegmann, A., Engelmann, N., & Willemsen, P. (2021). Can a Question Be a Lie? An Empirical Investigation. *Ergo*, *8*(0). https://doi.org/10.3998/ergo.1144

Walsh, K., & Currie, A. (2015). Caricatures, Myths, and White Lies. *Metaphilosophy*, *46*(3), 414–435. https://doi.org/10.1111/meta.12139

Waskan, J., Harmon, I., Horne, Z., Spino, J., & Clevenger, J. (2014). Explanatory Anti-Psychologism Overturned by Lay and Scientific Case Classifications. *Synthese*, *191*(5), 1013–1035. https://doi.org/10.1007/s11229-013-0304-2

Wiegmann, A., Samland, J., & Waldmann, M. R. (2016). Lying Despite Telling the Truth. *Cognition*, *150*, 37–42. https://doi.org/10.1016/j.cognition.2016.01.017

Wilkenfeld, D. A. (2013). Understanding as Representation Manipulability. *Synthese*, *190*(6), 997–1016. https://doi.org/10.1007/s11229-011-0055-x

Wilkenfeld, D. A. (2014). Functional Explaining: A New Approach to the Philosophy of Explanation. *Synthese*, *191*(14), 3367–3391. https://doi.org/10.1007/s11229-014-0452-z

Wilkenfeld, D. A., & Lombrozo, T. (2020). Explanation Classification Depends on Understanding: Extending the Epistemic Side-Effect Effect. *Synthese*, *197*(6), 2565–2592. https://doi.org/10.1007/s11229-018-1835-3

Wilkenfeld, D. A., Plunkett, D., & Lombrozo, T. (2018). Folk Attributions of Understanding: Is There a Role for Epistemic Luck? *Episteme*, *15*(1), 24–49. https://doi.org/10.1017/epi.2016.38

Williams, B. (2002). Truth and Truthfulness: An Essay in Genealogy. In *Truth and Truthfulness*. Princeton University Press. https://doi.org/10.1515/9781400825141

Zagzebski, L. (2001). Recovering Understanding. In M. Steup (Ed.), *Knowledge, Truth, and Duty* (pp. 235–252). Oxford University Press. https://doi.org/10.1093/0195128923.003.0015

Appendix

Remaining vignettes from Experiment 1

Evolution Story (positive intention)

Mary is a biology teacher at a summer camp for primary school students. During one of the sessions, she teaches the theory of evolution. She has some freedom in deciding how deep into the topic she will dive. One of the main claims of the theory of evolution is that humans share a common ancestor with great apes. Mary is aware that the theory of evolution is very complicated and decides to teach pupils a simplified version. She is a knowledgeable teacher so she could explain a more complicated and exact version of the theory. However, she presents a simplified version because she thinks that, at this level of education, it is pedagogically better to explain something that will be easier to understand. During her lesson, she says that human beings descended from apes, even though she knows that this statement is false. As a result of her presentation, the pupils acquire some understanding of evolution.

Evolution Story (negative intention)

Mary is a biology teacher at a summer camp for primary school students. During one of the sessions, she teaches the theory of evolution. She has some freedom in deciding how deep into the topic she will dive. One of the main claims of the theory of evolution is that humans share a common ancestor with great apes. Mary is aware that the theory of evolution is very complicated and decides to teach pupils a simplified version. She is a knowledgeable teacher so she could explain a more complicated and exact version of the theory. However, she presents a simplified version because she wants to deceive the pupils and create a wrong picture of the theory of evolution in front of them. During her lesson, she says that human beings descended from apes, even though she knows that this statement is false. As a result of her presentation, the pupils acquire some understanding of evolution.

Vignettes from Experiment 2

Solar System Story (positive intention)

Stan is a physics teacher at a summer camp for primary school students. One day, he teaches about our solar system. He has some freedom in deciding how he will explain this topic. He explains the heliocentric model, which says that planets move around the Sun in elliptical orbits. However, Stan decides to explain a simplified version of the heliocentric model because he believes that it is pedagogically better at this level of education. He does not want to overwhelm his pupils by giving them too much information. During his classes, he says that the planets in our solar system move around the Sun in circular orbits, even though he knows that this is a false depiction of our solar system. As a result of his lesson, the pupils acquire some understanding of our solar system.

Solar System Story (negative intention)

Stan is a physics teacher at a summer camp for primary school students. One day, he teaches about our solar system. He has some freedom in deciding how he will explain this topic. He explains the heliocentric model, which says that planets move around the Sun in elliptical orbits. However, Stan decides to explain a simplified version of the heliocentric model because he wants to deceive his pupils by giving them a false explanation of how our solar system is really built. During his classes, he says that the planets in our solar system move around the Sun in circular orbits, even though he knows that this is a false depiction of our solar system. As a result of his lesson, the pupils acquire some understanding of our solar system.

Laws of Motion Story (positive intention)

Jane is a physics teacher at a summer camp for primary school students. One of the topics concerns the laws of motion. It is up to her how deep into the topic she will dive. During her lessons, she explains the laws of motion on the basis of Newton's laws which are simple and relatively accurate for everyday use. She knows that Newton’s laws are false and that we have much more accurate theories at our disposal, like Einstein’s theories of relativity. Jane decides to explain the laws of motion using Newton’s laws because she thinks that her pupils would not understand an explanation of a more difficult theory. As a result of the presentation, her pupils acquire some understanding of the laws of motion.

Laws of Motion Story (negative intention)

Jane is a physics teacher at a summer camp for primary school students. One of the topics concerns the laws of motion. It is up to her how deep into the topic she will dive. During her lessons, she explains the laws of motion on the basis of Newton's laws which are simple and relatively accurate for everyday use. She knows that Newton’s laws are false and that we have much more accurate theories at our disposal, like Einstein’s theories of relativity. Jane decides to explain the laws of motion using Newton’s laws because she wants to deceive the students by giving them a false explanation. As a result of the presentation, her pupils acquire some understanding of the laws of motion.

Vignettes from Experiment 3

(Variants had only different endings.)

Laws of Motion

Jane is a physics teacher at a summer camp for primary school students. One of the topics concerns the laws of motion. It is up to her how deep into the topic she will dive. During her lessons, she explains the laws of motion on the basis of Newton's laws which are simple and relatively accurate for everyday use. She knows that Newton’s laws are false and that we have much more accurate theories at our disposal, like Einstein’s theories of relativity.

(neutral intention)

As a result of the presentation, her pupils acquire some understanding of the laws of motion.

(positive intention)

Jane decides to explain the laws of motion using Newton’s laws because she thinks that, at this level of education, it is pedagogically better to explain something that will be easier to understand. As a result of the presentation, her pupils acquire some understanding of the laws of motion.

(negative intention)

Jane decides to explain the laws of motion using Newton’s laws because she wants to deceive the students by giving them a false explanation. As a result of the presentation, her pupils acquire some understanding of the laws of motion.

Atom Story

John is a physics teacher at a summer camp for primary school students. One day, he teaches about the structure of an atom. He has some freedom in deciding how he will explain this topic. He explains the Bohr model of an atom, according to which electrons travel in circular orbits around the nucleus. John knows that this is a false depiction of an atom. He is aware that there are other, more exact but also much more complicated models of an atom.

(neutral intention)

As a result of the presentation, his pupils acquire some understanding of the structure of an atom.

(positive intention)

He presents the Bohr model because he thinks that, at this level of education, it is pedagogically better to explain something that will be easier to understand. As a result of the presentation, his pupils acquire some understanding of the structure of an atom.

(negative intention)

He presents the Bohr model because he wants to deceive his pupils about the real structure of an atom. As a result of the presentation, his pupils acquire some understanding of the structure of an atom.