Scaling up interactive argumentation by providing counterarguments with a chatbot

Sacha Altay [1 ✉], Marlène Schwartz[1], Anne-Sophie Hacquin[1], Aurélien Allard [2], Stefaan Blancke [3] and Hugo Mercier [1 ✉]

Discussion is more convincing than standard, unidirectional messaging, but its interactive nature makes it difficult to scale up. We created a chatbot to emulate the most important traits of discussion. A simple argument pointing out the existence of a scientific consensus on the safety of genetically modified organisms (GMOs) already led to more positive attitudes towards GMOs, compared with a control message. Providing participants with good arguments rebutting the most common counterarguments against GMOs led to much more positive attitudes towards GMOs, whether the participants could immediately see all the arguments or could select the most relevant arguments in a chatbot. Participants holding the most negative attitudes displayed more attitude change in favour of GMOs. Participants updated their beliefs when presented with good arguments, but we found no evidence that an interactive chatbot proves more persuasive than a list of arguments and counterarguments.

In many domains—from the safety of vaccination to the reality of anthropogenic climate change—there is a gap between the scientific consensus and public opinion[1]. The persistence of this gap in spite of numerous information campaigns shows how hard it is to bridge. It has even been suggested that information campaigns backfire, either by addressing audiences with strong pre-existing views[2,3] or by attempting to present too many arguments[4,5].

Fortunately, it appears that, in most cases, good arguments do change people's mind in the expected direction[6,7]. Still, the effects of short arguments aimed at large and diverse audiences, even if they are positive, are typically small[8–10]. By contrast, when people can exchange arguments face to face, more ample changes of mind regularly occur. Compare, for example, how people react to simple logical arguments. On the one hand, when participants are provided with a good argument for the correct answer to a logical problem, a substantial minority fails to change their mind[11,12]. On the other hand, when participants tackle the same problems in groups, nearly everyone discussing with a participant defending the correct answer changes their mind[11–13]. More generally, argumentation has been shown, in a variety of domains, to allow people to change their minds and adopt the best answers available in the group[13–17]. Even on contested issues, discussions with politicians[18], canvassers[19] or scientists[20,21] can lead to changes of mind that are significant, durable[19] and larger than those observed with standard messages[18,21].

Mercier and Sperber[17] have suggested that the power of interactive argumentation, in contrast to the presentation of a simple argument, to change minds stems largely from the opportunity that discussion affords to address the discussants' counterarguments. In the course of a conversation, people can raise counterarguments as they wish, the counterarguments can be rebutted, the rebuttals contested and so forth[22]. When people are presented with challenging arguments in a one-sided manner, as in typical messaging campaigns, they also generate counterarguments.[23–25] However, these counterarguments remain unaddressed. Arguably, the production of counterarguments that remain unaddressed is not only why standard information campaigns are not very effective but also why they sometimes backfire[26].

In a discussion, not only can all counterarguments be potentially addressed but only the relevant counterarguments are addressed. Different people have different reasons to disagree with any given argument. Attempting to address all the existing counterarguments should lead to the production of many irrelevant rebuttals, potentially diluting the efficacy of the relevant rebuttals. This might be why, in a normal conversation, we do not attempt to lay out all the arguments for our side immediately, waiting instead for our interlocutor's feedback to select the most relevant counterarguments to address[27].

Unfortunately, discussion does not scale up well—indeed, it is most natural in groups of at most five people[28,29]. Here, we developed and tested two ways of scaling up discussion. The first consisted in gathering the most common counterarguments for a given issue and a given population and creating a message that rebuts the most common counterarguments as well as the responses to the rebuttals, as would happen in a conversation. The issue, then, is that many—potentially most—of these counterarguments are likely to be irrelevant for most of the audience. As a result, we developed a second way of scaling up discussion by using a chatbot in which participants could select which counterarguments they endorse and only see the rebuttals to these counterarguments. Studies on argumentation using chatbots or similar automated computer-based conversational agents suggest that they can be useful to change people's mind[30,31] and that asking users what they are concerned about increases chatbots' efficacy by providing users with more relevant counterarguments[32].

In the remaining of the introduction, we present the topic we have chosen to test our methods for scaling up discussion, as well as the design of the experiment, and how the different conditions were constructed. Finally, specific hypotheses are introduced.

We choose genetically modified organisms (GMOs) and genetically modified (GM) food as a topic for our experiment because, despite the broad scientific consensus on GM food safety for human health[33–39], public opinion remains, in many countries, staunchly opposed to GM food and GMOs more generally[40–43]. In the United States, it is the topic on which discrepancy between

[1]Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, Paris, France. [2]University of Geneva, Geneva, Switzerland. [3]Department of Philosophy, Tilburg University, Tilburg, the Netherlands. ✉e-mail: sacha.altay@gmail.com; hugo.mercier@gmail.com

579

scientists and laypeople's opinion is highest[1]. In France, where the pilot study was conducted (see Pilot data section and Pilot study section in Supplementary information), rejection of GMO is pervasive.[40] Indeed, 84% of the public thinks that GM food is highly or moderately dangerous[44], and 79% of the population is worried that some GMO may be present in their daily diet[45]. In the United Kingdom, where the pre-registered study was conducted, rejection of GMO is common. There, 45% of the public thinks that GM food is dangerous (while only 25% think that it is not)[40], and 58% of the public does not want to eat this type of food (while only 24% wants to)[40]. On the whole, British people appear to be largely un-persuaded by the benefits of GMOs[46–48]. The gap between the scientific consensus and public opinion on GMOs is all the more problematic since GM food and GMOs more generally can not only improve health and food security but also help fight climate change[37,49,50].

Our goal was thus to test whether rebutting participants' counterarguments against GMOs will lead them to change their mind on this topic. To properly evaluate the efficiency of this intervention, we used the following four conditions.

First, as a control condition, we provided participants with a sentence describing what GMOs are. Given that no persuasion should take place in this condition, any attitude change (measured as the difference between the pre- and post-intervention attitudes) would reflect task demands and can thus be used as a baseline against which to compare attitude change in the other conditions.

Second, we compared our interventions with one of the most common techniques used to bridge the gap between scientific consensus and public opinion, that is, informing the public of the existence and strength of the scientific consensus (the so-called gateway belief model). Some studies using this gateway belief model have proven it to be effective at reducing the gap between public opinion and the scientific consensus on a variety of topics[51–56] (although see refs. [8,9]). This consensus condition allowed us to tell whether our interventions could improve attitude change by comparison with a popular messaging strategy.

Third, in the counterarguments condition, participants were provided with a series of counterarguments against GMOs, rebuttals against these counterarguments, counterarguments of these rebuttals and so forth (for at most four steps; how these arguments were created is described in the Design section). One of these counterarguments mentions the existence and strength of the scientific consensus, as in the consensus condition. Comparing the attitude change obtained in the consensus and counterarguments conditions allowed us to test whether countering participants' arguments, instead of only presenting a forceful argument, was more effective at changing people's mind.

Fourth, in the chatbot condition, participants could read exactly the same materials as in the counterarguments condition, but through a chatbot, enabling them to easily access the most relevant, and only the most relevant, rebuttals to their counterarguments (the workings of the chatbot are detailed in the Design section; see Fig. 1 for a visualization of the chatbot's interface). Comparing the changes of minds obtained in the chatbot and the counterarguments condition allowed us to test whether presenting participants only with the rebuttals that are most relevant for them leads to more ample changes of mind.

Comparison of these four conditions allowed us to tell whether (i) any of these interventions resulted in attitude change, (ii) whether the attitude change was larger when arguments were provided (that is in the consensus, counterarguments and chatbot conditions), (iii) whether any argument-driven attitude change was larger when rebuttals to counterarguments were provided (counterarguments and chatbot conditions) and (iv) whether any rebuttal-driven attitude change was larger when only relevant rebuttals were provided (chatbot condition).

On the basis of the literature reviewed above, we derived the following hypotheses. First, the literature on the gateway belief model, on the importance of addressing counterarguments and on the importance of only addressing relevant counterarguments, led to the following hypotheses:

H1: Participants will hold more positive attitudes towards GMOs after the experimental task in the consensus condition than in the control condition, controlling for their initial attitudes towards GMOs.

H2: Participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the control condition, controlling for their initial attitudes towards GMOs.

H3: Participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the control condition, controlling for their initial attitudes towards GMOs.

H4: Participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the consensus condition, controlling for their initial attitudes towards GMOs.

H5: Participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the counterarguments condition, controlling for their initial attitudes towards GMOs.

H6: In the chatbot condition, the number of arguments explored by participants will predict holding more positive attitudes towards GMOs after the experimental task, controlling for their initial attitudes towards GMOs (that is, exploring more arguments should lead to more positive attitude change).

H7: In the chatbot condition, time spent on the task should lead to more positive attitudes towards GMOs after the experimental task than time spent on the counterarguments condition, controlling for initial attitudes towards GMOs.

Participants were given the opportunity to read many more arguments in the counterarguments condition and in the chatbot condition than in the consensus condition. Models of attitude change—such as the elaboration likelihood model[57]—suggest that participants might use the number of arguments as a low-level cue that they should change their minds, at least when they are not motivated to process the arguments in any depth[58]. However, it has also been argued that presenting people with too many arguments—even good ones—might make a message less persuasive if the misinformation that the arguments aim to correct is simpler and more appealing[5], so that more is not necessarily best when it comes to the number of arguments provided. Still, if H6 and H7 proved true, it could be argued that participants use a low-level heuristic in which they are convinced by the sheer number of arguments instead of the content of the arguments. If people use the number of arguments in this manner, it should affect their overall attitudes towards GMOs. By contrast, if people pay attention to the content of the arguments, the arguments should only change the participants' mind on the specific topic they bear upon, leading us to the following hypothesis:

H8: In the chatbot condition, participants will hold more positive attitudes after the experimental task on issues for which they have explored more of the rebuttals related to the issue, controlling for their initial attitudes on the issues, the type of issue and the total (that is, related to the issue or not) number of arguments explored.

Finally, given that the backfire effect has been observed in several experiments[2,4,59–61] but has rarely, or not at all, been observed in several large-scale studies[6,7,62–64], we formulated the following hypothesis:

H9: H1, H2 and H3 also hold true among the third of the participants initially holding the most negative attitudes about GMOs. (Note that this criterion is more stringent than an absence of backfire effect, as it claims that there will be a positive effect even among participants with the most negative initial attitudes.)
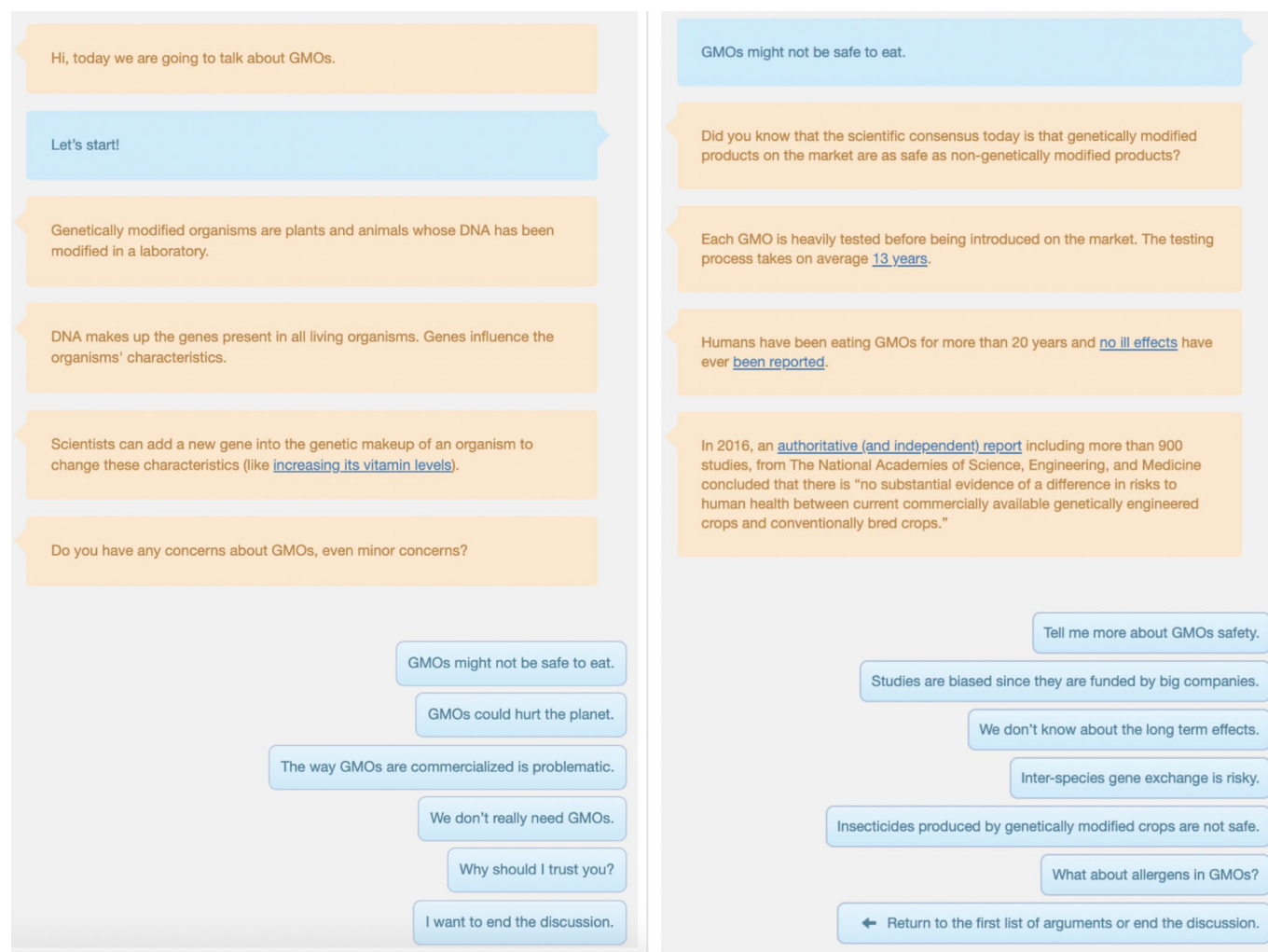
Fig. 1 | The beginning of a conversation with the chatbot. The screen-wide dialogue bubbles correspond to past interactions including the participant's counterarguments (blue) and the chatbot's rebuttals (beige) as well as the participant's choices at this stage of the interaction (right-justified blue bubbles).

Although it is methodologically impossible to completely disentangle the effects of the mode of presentation (for example, degree of interactivity) and of the specific information presented, the present experiment provides a test of whether addressing people's counterarguments, in particular by using an interactive chatbot, results in attitude changes that are larger than those obtained with a common messaging technique. From a theoretical point of view, these results help us better understand the process of attitude change, potentially highlighting its rationality. If people are sensitive to the rebuttals of their counterarguments, it suggests that their rejection of the initial argument was not driven by sheer pigheadedness but by having unanswered counterarguments. From an applied point of view, positive results would provide an efficient and easy-to-use tool to help science communicators bridge the gap between scientific consensus and public opinion.

Results

In the control condition, participants read a sentence describing what GMOs are. In the consensus condition, they read a paragraph on the scientific consensus on the safety of GMOs. In the counterarguments condition, they were exposed to the most common counterarguments against GMOs, together with their rebuttal. In the chatbot condition, they were exposed to the same arguments as in the counterarguments condition but through a chatbot (that

is, instead of scrolling, they had to click to make the arguments appear). The effect of the treatments on participants' attitudes towards GMOs is depicted in Figs. 2 and 3.

Confirmatory analyses. In line with H1, participants held more positive attitudes towards GMOs after the treatment in the consensus condition than in the control condition ($b = 0.37$ [0.23, 0.51], $t(1,144) = 5.36$, $P < 0.001$, Bayes factor ($BF_{10}) = 10^4$). In other words, being exposed to the scientific consensus led to a 0.37 point (on a 7-point scale) move towards a pro-GMO opinion.

In line with H2, participants held more positive attitudes towards GMOs after the treatment in the counterarguments condition than in the control condition ($b = 0.99$ [0.86, 1.12], $t(1,144) = 14.65$, $P < 0.001$, $BF_{10} = 10^{41}$). That is, scrolling through the chatbot's arguments led to a 0.99-point move towards pro-GMO opinion.

In line with H3, participants held more positive attitudes towards GMOs after the treatment in the chatbot condition than in the control condition ($b = 0.77$ [0.63, 0.90], $t(1,144) = 11.38$, $P < 0.001$, $BF_{10} = 10^{25}$). In sum, interacting with the chatbot by clicking on the arguments led to 0.77-point move towards pro-GMO opinion.

In line with H4, participants held more positive attitudes towards GMOs after the treatment in the counterarguments condition than in the consensus condition ($b = 0.62$ [0.49, 0.75], $t(1,144) = 9.15$, $P < 0.001$, $BF_{10} = 10^{16}$). In other words, scrolling through the chatbot's
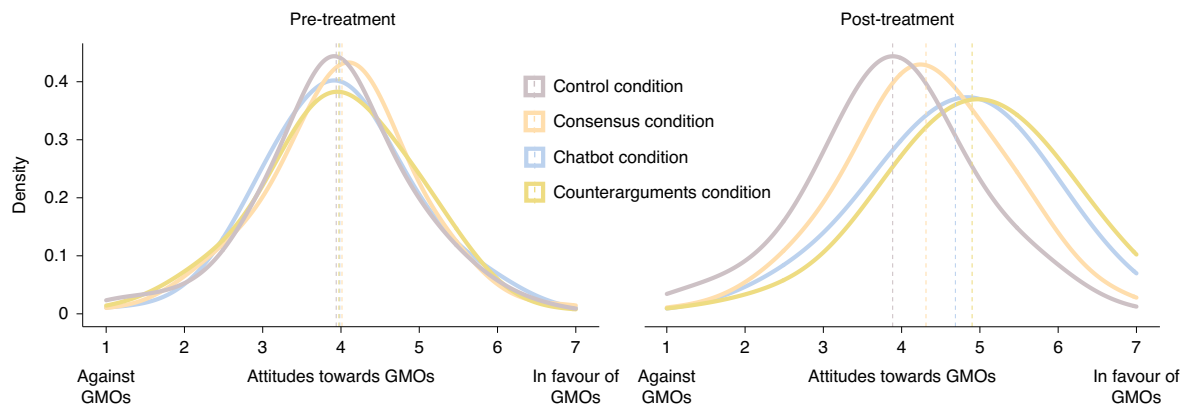
Fig. 2 | Distribution of participants' attitudes towards GMOs. Density plots representing the distributions of participants' attitudes towards GMOs before (left) and after treatment (right) in the four conditions. Dashed lines indicate the means in each condition. Control condition ($N = 275$): a sentence describing what GMOs are. Consensus condition ($N = 273$): a paragraph on the scientific consensus on GMOs safety. Counterarguments condition ($N = 299$): a text with the most common counterarguments against GMOs, together with their rebuttal. Chatbot condition ($N = 302$): the same arguments as in the counterarguments condition but accessed interactively via a chatbot.

arguments led to a 0.62-point move towards pro-GMO opinion compared with exposure to the scientific consensus.

Contrary to H5, participants held more positive attitudes towards GMOs after the treatment in the counterarguments condition than in the chatbot condition ($b = 0.22$ [0.09, 0.35], $t(1,144) = 3.37$, $P < 0.001$, $BF_{10} = 15.32$). That is, scrolling through the chatbot's arguments led to a 0.22-point move towards pro-GMO opinion compared with interacting with the chatbot by clicking on the arguments.

In line with H6, the number of arguments explored by participants in the chatbot condition predicted holding more positive attitudes towards GMOs after the treatment ($b = 0.04$ [0.02, 0.06], $t(299) = 4.09$, $P < 0.001$, $BF_{10} = 14.41$). That is, each additional argument that participants explored led to a 0.04-point move towards pro-GMO opinion.

Contrary to H7, time spent in the chatbot condition did not lead to significantly more positive attitudes towards GMOs after the treatment than time spent on the counterarguments condition ($b = 0.004$ [−0.05, 0.04], $t(596) = 0.20$, $P = 0.84$, $BF_{10} = 0.08$). This effect is negligible as the 90% CI [−0.08, 0.06] falls inside the pre-registered [−0.1, 0.1] interval corresponding to an effect smaller than $\beta = 0.1$. Figure 4 offers a visual representation of the interaction. In both conditions, time spent on the task led to more positive attitudes towards GMOs ($b = 0.07$ [0.05, 0.09], $t(596) = 7.54$, $P < 0.001$, $BF_{10} = 10^{12}$; chatbot condition: $b = 0.07$ [0.03, 0.10], $t(299) = 3.36$, $P < 0.001$, $BF_{10} = 19$; counterarguments condition: $b = 0.07$ [0.05, 0.09], $t(296) = 7.35$, $P < 0.001$, $BF_{10} = 10^9$). In other words, the longer participants spent on the chatbot's interface, the more they developed pro-GMO opinions.

Contrary to H8, participants did not hold significantly more positive attitudes after the treatment on issues for which they had explored more of the related rebuttals ($b = 0.006$ [−0.02, 0.04], $t(898.9) = 0.09$, $P = 0.77$, $BF_{10} = 0.09$). This effect is negligible as the 90% CI [−0.04, 0.07] falls inside the pre-registered [−0.1, 0.1] interval corresponding to an effect smaller than $\beta = 0.1$. For instance, participants exploring more arguments related to GM food safety did not develop more pro-GMO opinions on GM food safety compared with other issues.

In line with H9, H1–3 held true among the third of the participants initially holding the most negative attitudes towards GMOs (chatbot condition: $b = 0.93$ [0.67, 1.20], $t(376) = 6.92$, $P < 0.001$, $BF_{10} = 10^8$; counterarguments condition: $b = 1.35$ [1.09, 1.62], $t(376) = 9.96$, $P < 0.001$, $BF_{10} = 10^{17}$; consensus condition: $b = 0.47$

[0.29, 0.75], $t(376) = 3.33$, $P < 0.001$, $BF_{10} = 24$). That is, being exposed to the consensus, and scrolling or clicking through the chatbot's arguments, led to more pro-GMO opinions even among the third of participants initially holding the most anti-GMO opinions.

Exploratory analyses. To assess whether time spent on the task might explain the greater impact of the counterarguments condition compared with the chatbot condition, we tested whether participants had spent more time in the former than the latter, revealing that they had (counterarguments condition, $M = 5.74$ min, s.d. 5.63 min; chatbot condition, $M = 3.70$ min, s.d. 2.56 min; $t(415.48) = 5.73$, $P < 0.001$). In a regression model without time as predictor, condition (chatbot versus counterarguments) was a significant predictor of attitude change ($b = 0.23$ [0.07, 0.38], $t(599) = 2.83$, $P = 0.008$), but when adding time spent in the model, the effect of condition was not significant anymore ($b = 0.08$ [−0.07, 0.23], $t(598) = 1.01$, $P = 0.37$), whereas the effect of time was ($b = 0.07$ [0.05, 0.09], $t(598) = 8.29$, $P < 0.001$). Then, a mediation analysis (non-parametric bootstrap confidence intervals with the percentile method[65]) suggested that 65% of the effect of condition was mediated by time ([0.37, 1.84], $P = 0.004$), with an indirect effect via the time mediator estimated to be $b = 0.15$ [0.10, 0.21] ($P < 0.001$). However, this should not be taken as proof of causality because non-observed variables could create (or inflate) the correlation observed between time and attitude change[66]. Nevertheless, time remains a credible mediator since it plausibly plays a role in attitude change, and more time spent reading the arguments might translate into greater attitude change.

To investigate H9 further, we examined the relationship between participants' initial attitudes, and attitude change. More precisely we tested the interaction between participants' initial attitudes and the experimental condition (with the control condition as baseline) on attitude change. By contrast with H9, here all the participants are included in the analysis. We found that, compared with the control condition, participants initially holding more negative attitudes displayed more attitude change in favour of GMOs in the counterarguments condition ($b = 0.30$ [0.17, 0.44], $t(1,144) = 4.47$, $P < 0.001$) and in the chatbot condition ($b = 0.19$ [0.06, 0.33], $t(1,144) = 2.81$, $P = 0.008$), but only marginally in the consensus condition ($b = 0.14$ [0.006, 0.28], $t(1,144) = 2.05$, $P = 0.06$).

We also found that H1–3 held true for each question of the GMOs attitudes scale: participants deemed GM food to be safer to eat, and less bad for the environment, reported being less worried
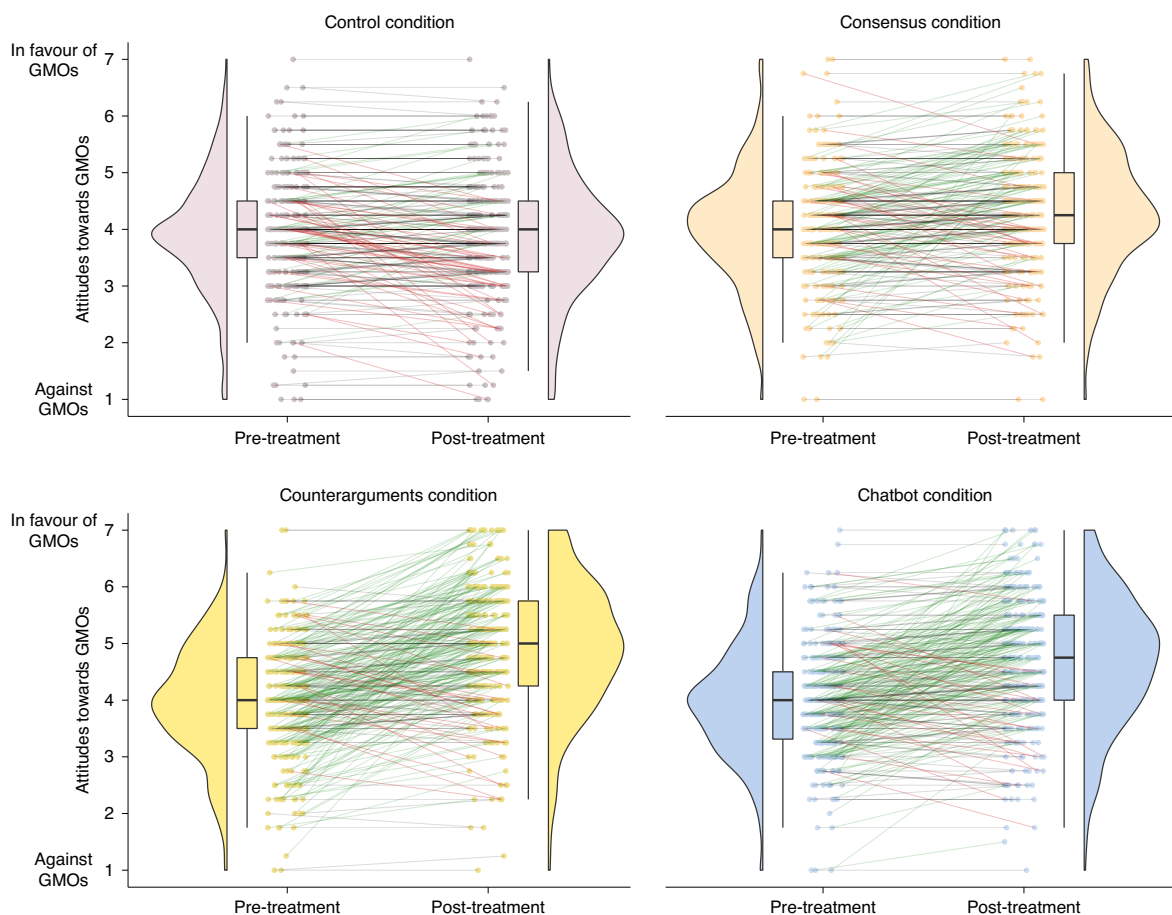
Fig. 3 | Evolution of participants' attitudes toward GMOs. Each dot represents a participant, and the lines connecting the dots represent the change in each participant's attitudes before and after treatment. Grey lines represent participants whose attitude toward GMOs was similar after and before the treatment (that is, on a four-point Likert scale, their attitude did not change by more than one point overall). Among the other participants, green (red) lines represent participants whose attitude toward GMOs was more positive (negative) after the treatment than before. In the box plots, the boxes represent the middle 50% of the scores (the boundaries of the boxes represent the 25th and 75th percentile), the lines dividing the boxes are the median and the boundaries of the upper and lower whiskers were determined by multiplying the inter-quartile range by 1.5. Density plots represent the distributions of participants' attitudes towards GMOs before and after treatment. Control condition ($N=275$): a sentence describing what GMOs are. Consensus condition ($N=273$): a paragraph on the scientific consensus on GMOs safety. Counterarguments condition ($N=299$): a text with the most common counterarguments against GMOs, together with their rebuttal. Chatbot condition ($N=302$): the same arguments as in the counterarguments condition but accessed interactively via a chatbot.

about the socio-economic impacts of GMOs and perceived GMOs as more useful after the treatment in the consensus condition, counterarguments condition and chatbot condition compared with the control condition (see Exploratory analyses section in Supplementary information).

Finally, for three out of the four main arguments on GMOs, the best predictor of whether a participant clicked on a given argument in the chatbot was how negative their initial attitudes were regarding that argument (see Exploratory analyses section in Supplementary information). This suggests that participants selected arguments that addressed their concerns, instead of arguments that might have reinforced their priors.

Discussion

We investigated whether addressing many of participants' arguments against GMOs would result in significant changes of mind on that issue. First, despite previous failures to apply the gateway belief model to GMOs[8,9], we found that a simple argument pointing out the existence of a scientific consensus on the safety of GMOs led to more positive attitudes towards GMOs ($\beta = 0.32$). Second, we found that

addressing many of the participants' arguments against GMOs led to much more positive attitudes towards GMOs (counterarguments condition, $\beta = 0.85$; chatbot condition, $\beta = 0.66$). These effect sizes compare very favourably with those observed in other interventions, such as refs. [8–10,62,67–69]. After reading the rebuttals against criticisms of GMOs, a large number of participants adopted strongly pro-GMOs views: the number of participants with an average score of at least 5 (on the 1–7 attitude scale) went from 104 to 299 (out of 601), and the number of participants with an average score of at least 6 went from 15 to 107 (out of 601). A subsequent study found similar results in another domain: a chatbot aimed at alleviating concerns about coronavirus disease 2019 (COVID-19) vaccines generated more positive attitudes towards these vaccines, and higher intentions to vaccinate[69].

Our results reveal that participants changed their mind more as they spent more time reading counterarguments, and tended to spend more time when all the counterarguments were available (counterarguments condition) than when they were offered the possibility of only selecting the most relevant counterarguments (chatbot condition). Moreover, being exposed only to the counterarguments that participants had selected, by contrast with all the

Fig. 4 | Relationship between time spent on the treatment and attitude change in the counterarguments and chatbot conditions. The blue and yellow lines represent the regression lines of attitude change on time spent in each condition. The shaded area represents the 95% CI. The density plots represent the distributions of attitude change and time spent in each condition. Counterarguments condition ($N = 299$): a text with the most common counterarguments against GMOs, together with their rebuttal. Chatbot condition ($N = 302$): the same arguments as in the counterarguments condition but accessed interactively via a chatbot.

counterarguments, did not make the counterarguments more efficient. It is possible that participants used the sheer number of arguments presented as a cue to change their mind. It is also plausible that, in the case at hand, all the counterarguments presented to the participants were sufficiently relevant that none detracted from the persuasiveness of the whole set, or that participants selected the most relevant via scrolling, and that this selection was more efficient than via clicking. If this is the case, then the main reason for the increased efficiency of the chatbot (controlling for time spent), that is, that people avoid reading irrelevant arguments, disappears.

This finding has practical consequences: given the available evidence, it is probably best to give chatbot users the option to scroll through the arguments instead of clicking on them, as in our counterarguments condition. In the chatbot that we used to inform French people about COVID-19 vaccines, we gave users the possibility of scrolling through the arguments instead of clicking on them[69]. In that study, users selecting the non-interactive chatbot did so as a complement of the interactive chatbot.

In line with a growing body of literature[7,70], we found no evidence that participants initially holding more negative attitudes towards GMOs held even more negative attitudes towards GMOs after having been exposed to arguments in favour of GMOs. Instead, we found the opposite pattern: participants initially holding more negative attitudes displayed more attitude change in favour of GMOs. Similar evidence suggests that corrections work best on people who are the most misinformed[71–73] and that, in general, those whose attitudes were initially furthest from the facts changed their minds the most towards the facts[20,69].

These results are good news for science communicators, showing that participants can be convinced by good, well-supported arguments. Moreover, the initially very negative attitudes of some participants did not prove an obstacle to changing their mind. This should encourage science communicators to discuss heated topics with the public, even with those furthest away from the scientific consensus (see also refs. [20,62]).

Our study has some important limitations. First, we do not know the impact our chatbot would have in more naturalistic settings. In the present experiments, participants were paid to interact with the chatbot. How many people would spontaneously want to spend five minutes learning about GMOs without being paid? And would people opposed to GMOs be less likely to interact with it? Second, all our dependent variables are declarative, and attitudes do not always translate into behaviours. For instance, we do not know whether the chatbot will increase people's willingness to buy and consume GM products. Third, our sample of participants is not representative of the general population. The chatbot might work better on some segments of the population than others, such as young and educated people (although the experimental manipulations should be robust across samples, see ref. [74]). All these metrics are key to estimate the chatbot's effects outside of an experimental setting.

Exploratory hypotheses point to two interesting patterns in our data. First, participants' behaviour in the chatbot was in line with their attitudes, as they selected the issues on which they had the most negative attitudes, thereby exposing them to the most relevant counterarguments. Second, the counterarguments—including simply providing information about the scientific consensus—had effects beyond the specific issue they addressed. While this might suggest that participants were falling prey to a kind of halo effect, it is also possible that participants drew judicious inferences from one set of arguments to others. For example, participants who come to accept that GMOs are safe to eat might also see them as more useful.

At first glance, our results might seem to suggest that presenting counterarguments in a chatbot, by contrast with a more standard text, offers little advantage, or might even prove less persuasive. However, it should be noted that, even when not presented in a chatbot, the counterarguments were organized according to a clear dialogic structure (the exact same one as the chatbot), which might have facilitated their understanding and the identification of the most relevant counterarguments. Moreover, it is possible that participants not expressly paid to take part in an experiment might find the chatbot's interactivity more alluring than a standard text. Future experiments should investigate whether that is the case. Another promising avenue for future research is whether the very large effects observed here persist in time (see, for example, ref. [19]).

Methods

Ethics information. The present research received approval from an ethics committee (CER-Paris Descartes, no. 2019-03-MERCIER). Participants were presented with a consent form and had to give their informed consent to participate in the study. They were paid £1.38.

Pilot data. Among 147 French participants who pre-tested the chatbot, we found that:

- Participants' attitudes toward GMOs became more positive after having interacted with the chatbot ($t(69) = 3.68$, $P < 0.001$, $d = 0.28$, 95% CI [0.13, 0.44]).
- The number of arguments explored by participants significantly predicted a larger shift towards positive attitudes towards GMOs ($\beta = 0.23$, 95% CI [0.09, 0.37], $t(67) = 3.32$, $P = 0.001$).
- Participants who only provided their attitudes toward GMOs after having interacted with the chatbot did not have significantly different attitudes towards GMOs compared with participants who provided their attitudes toward GMOs both before and after having interacted with the chatbot ($t(138.57) = 0.71$, $P = 0.48$, $d = 0.14$, 95% CI [−0.21, 0.44]).
- Participants judged the bot as quite enjoyable ($M = 60.13$, s.d. 25.3), intuitive ($M = 65.94$, s.d. 26.45) and not very frustrating ($M = 35.39$, s.d. 31.22) (all scales from 0 to 100).

More details about the pilot can be found in Pilot study section in Supplementary information.

Design. To create the counterarguments condition, we systematically gathered the most common counterarguments to the acceptance of GMOs, relying on a variety of methods. First, we drew on popular anti-GMOs websites (such as nongmoproject.org), and on the scientific literature on public opinion towards GMOs[40,77–80]. Second, we relied on the expertise of two of the co-authors, who have both participated in public events about GMOs[20,81]. Third, we conducted

a preliminary study in which we asked participants to rate how convincing and how accurate they found our rebuttals to the most common counterarguments against GMOs. When the rebuttals were found to be unconvincing, participants were asked to explain what made the rebuttals unconvincing and write any counterarguments that came to their mind that could weaken the rebuttals (participants who found the rebuttals convincing were also asked to explain why they found them convincing). At the end of the preliminary study, participants were asked to write whether they had any counterarguments against GMOs that had not been raised during the experiment. This ensured that we covered most of the arguments people hold against GMOs and that the rebuttals to these counterarguments were taken seriously.

To develop the rebuttals to the most common counterarguments, we relied on personal communication with an expert on GMOs, the website gmoanswers.com, the scientific literature on attitudes towards GMOs[40,77–80], the scientific literature on GMOs[36,82–85] and Wikipedia as well as the publications of scientific agencies[33–35].

The counterarguments and rebuttals were used to build the counterarguments condition. In this condition, participants are presented on the chatbot interface with the counterarguments and rebuttals available on the chatbot. However, participants cannot select the arguments. Instead, they have to scroll to read the counterarguments and rebuttals. The only difference between the counterarguments condition and the chatbot condition is the interactivity of the chatbot (that is, having to click on the counterarguments, seeing the rebuttals appear progressively instead of instantly and having the option of not displaying at all some rebuttals). We estimated the reading time for the counterarguments and rebuttals (~3,000 words) to be approximately 11 min (for a reading time of four words per second[86]).

In the counterarguments condition, participants were exposed to the most common counterarguments that we gathered against GMOs, as well as the rebuttals of these counterarguments. However, many participants might not share the concerns expressed in some counterarguments and thus find the rebuttals largely irrelevant. To address this problem, we created a chatbot whose content was identical to the content of the counterarguments condition but in which participants had to select (by clicking on them) the counterarguments against GMOs (or against the rebuttals to their previous counterarguments) that they were most concerned about, and they were provided with rebuttals addressing the selected counterargument.

The chatbot was organized as follows. After a brief technical description of GMOs (used in part in the control condition), participants were asked whether they had any concerns about GMOs and were given a choice of four counterarguments to select from: 'GMOs might not be safe to eat', 'GMOs could hurt the planet', 'The way GMOs are commercialized is problematic', and 'We don't really need GMOs'. Participants were also be able to select, at any stage, an option 'Why should I trust you?', which informed them about who we were, who funded us and what our goals were (all the materials are available on the Open Science Framework (OSF) at https://osf.io/cb7wf/).

Each time participants selected a counterargument, the chatbot offered a rebuttal. Participants could select between several counterarguments to these rebuttals, which were addressed by the chatbot as they were selected. In total, the chatbot offered 35 counterarguments against GMOs, together with their 35 rebuttals. Participants were not able to write open-ended counterarguments addressed to the chatbot. They were only able to select among the counterarguments offered, to which the chatbot answered with a predefined rebuttal. If the rebuttal exceeded five lines, it was displayed in separate discussion bubbles appearing progressively to give participants the impression that the bot was typing. As an example, here is the text participants saw after selecting the first counterargument that the chatbot presented at each step (sections in brackets did not appear to the participants):

Participant [first counterargument]: GMOs might not be safe to eat.

Chatbot [first rebuttal]: Did you know that the scientific consensus today is that genetically modified products on the market are as safe as non-genetically modified products? Each GMO is heavily tested before being introduced on the market. The testing process takes on average 13 years. Humans have been eating GMOs for more than 20 years and no ill effects have ever been reported. In 2016, an authoritative (and independent) report including more than 900 studies, from The National Academies of Science, Engineering, and Medicine concluded that there is "no substantial evidence of a difference in risks to human health between current commercially available genetically engineered crops and conventionally bred crops."

Participant [follow-up counterargument]: We don't know about the long-term effects.

Chatbot [follow-up rebuttal]: After **over 40 years** of research we have a good idea of the long-term effects of genetically modified food. On genetically modified corn alone more than 6000 studies have been published in scientific journals. A recent independent review of the scientific literature on long-term effects of genetically modified food concluded that: "genetically modified plants are nutritionally equivalent to their non-genetically modified counterparts and can be safely used in food and feed."

Arguments in favour of GMOs contained hyperlinks to scientific articles, reports from scientific agencies, and Wikipedia pages (which were identical

in the counterarguments condition). At any time, users had the possibility of coming back to the first four basic counterarguments of the main menu, or of exiting the chatbot.

Experimental procedure. Participants were asked to either read a simple explanation of what a GMO is (control condition), read a short paragraph on the scientific consensus on the safety of GM food (consensus condition), read counterarguments to GMOs accompanied by rebuttals of these arguments (counterarguments condition) or explore the same counterarguments and rebuttals by interacting with a chatbot (chatbot condition) (see Table 1).

Participants first had to complete a consent form and answer a few questions to measure their attitudes towards GMOs. Participants had to express the extent to which they agreed with the following four statements on a seven-point Likert scale:

- Genetically modified food is **safe** to eat.
- Genetically modified organisms (GMO) are **bad** for the environment.
- GMOs are **useless**.
- I'm worried about the socio-economic impacts of GMOs (on farmers in poor countries, wealth distribution, lack of competition, etc.).

In all analyses (except for H8), these four variables were treated as a single composite variable that we refer to as 'GMOs attitude'. Next, participants were presented with one of the four following conditions:

- Control condition
- Consensus condition
- Counterarguments condition
- Chatbot condition

Participants were randomly assigned to one of the four conditions by a pseudo-randomizer on the survey platform Qualtrics (that is, a randomizer that ensures that an equal number of participants is attributed to each condition). In all the conditions, participants were told to spend as much or as little time as they wanted interacting with the chatbot and exploring the text. By doing so, we improved the ecological validity of the task as participants were explicitly given leeway to engage with the arguments to the extent they wished—as they would if they had encountered the arguments in any other setting. Once they finished reading the arguments, participants answered the same questions regarding their GMOs attitudes as before the experimental task. Finally, participants provided basic demographic information (age, gender and education). Since data collection was automatized on Qualtrics and Prolific Academic, all our statistical analyses were pre-registered and there was no subjective coding of the data, the experimenters were not blind to the conditions of the experiments. Participants were not blinded to the study hypotheses. However, since the experiment had a between-participants design and most of our hypotheses (except H6 and H8) bear on comparisons across conditions, participants should not have been able to infer our hypotheses and act accordingly. See Table 1 for an overview of our design.

Materials. The neutral GMOs description used in the control condition reads as follows:

Genetically modified organisms are plants and animals whose DNA has been modified in a laboratory.

In the consensus condition, participants were provided with an account of the scientific consensus accompanied by sources. This account was more detailed than the ones used by most studies highlighting the scientific consensus on GMOs, such as 'Did you know? A recent survey shows that 90% of scientists believe genetically modified foods are safe to eat.'[9] The text used in the present experiment was:

There is a scientific consensus on the fact that genetically modified products on the market are as safe as non-genetically modified products. In 2016, an authoritative (and independent) report including more than 900 studies, from The National Academies of Science, Engineering, and Medicine concluded that there is "no substantial evidence of a difference in risks to human health between current commercially available genetically engineered crops and conventionally bred crops." 88% of scientists of the American Association for the Advancement of Science think that GM crops are safe to eat.

All the materials can be found on OSF at https://osf.io/cb7wf/ (in French and in English). The control and the consensus condition were displayed on the survey platform Qualtrics. The chatbot and the counterarguments condition (composed of all the counterarguments and rebuttals available on the chatbot) were displayed on the same, custom-made website. The only difference between the two conditions were that, in the chatbot condition, participants selected counterarguments and thus only saw the rebuttals that address these counterarguments, whereas in the counterarguments condition, participants scrolled through all the counterarguments and rebuttals. Figure 1 offers a visualization of the chatbot's interface.

Sampling plan. Based on the literature and on our pilot study (see below), we expected the effect of the chatbot on the evolution of attitudes towards GMOs to be large. Previous studies have shown that learning about the science behind genetic modification technology leads to more positive attitudes towards GMOs (ANOVA, $P < 0.001$, $\eta^2 = 0.09$)[67], as does discussion of the scientific evidence on GMOs safety in small groups (ANOVA, $P < 0.001$, $\eta^2 = 0.45$)[20]. In our pilot study, we found

Table 1 | Design table

	Hypothesis	Sampling (for example, power analysis)	Analysis	Interpretation given to different outcomes
H1	Participants will hold more positive attitudes towards GMOs after the experimental task in the consensus condition than in the control condition, controlling for their initial attitudes towards GMOs.	We performed an a priori power analysis with G*Power3 (ref. [75]). To compute the necessary number of participants, we decided that the minimal effect size of interest would correspond to a Cohen's d of 0.2 between two different experimental conditions, since this corresponds to what is generally seen as a small effect[76]. Based on a correlation of 0.75 between initial and final GMOs attitudes (estimated from the pilot), we need at least 275 participants per condition to detect this effect, at an α level of 5% and a power of 95% and based on a two-tailed test (see Supplementary information for more details). It is expected that approximately 15% of participants may encounter problems accessing the chatbot's interface (a percentage estimated while pre-testing the chatbot). We excluded participants who were not able to access the chatbot's interface. To anticipate these losses, we recruited 324 participants (275/0.85) instead of 275 in the chatbot condition and in the counterarguments condition. A total of 1,198 United Kingdom participants were recruited on the crowd-sourcing platform Prolific Academic. Data collection stopped when each condition reached the minimum number of participants required by the power analysis after exclusions. That is, if 30% of participants encountered problems accessing the chatbot's interface, we recruited 30% additional participants in the conditions where it will be needed. During the stage 2 peer-review process, reviewers noticed that our a priori power analysis at stage 1 had not taken into account our correction for multiple testing. In addition, we considered that, to compute a difference in mean between two experimental conditions, adding participants to unrelated conditions does not affect power. To enable readers to evaluate the impact of this on our power calculations, we repeated our simulations with fixed sample sizes and a fixed effect size (the a priori defined minimum effect size of interest, that is, $d = 0.2$). These simulations showed that actual power to detect the smallest effect size of theoretical interest was higher than 91% in all conditions (see Actual power section in Supplementary information; the R code associated with the simulations is available on OSF in the 'Stage 2 Registered Report' section (https://osf.io/cb7wf/)).	H1–3 are tested on the full dataset with one multivariate regression. Attitudes after the experimental task are set as the dependent variable, while attitudes before the experimental task and condition are set as predictors. The control condition is set as the baseline for the variable condition. In other words, consensus condition, counterarguments condition and chatbot condition are each compared with control condition. Attitudes before the experimental task are mean-centred to facilitate interpretation of the intercept, which corresponds to the mean post-attitude for the control condition.	For all interpretations, the effects are characterized not only by their statistical significance (below the stated α threshold of 0.05) but also by their size. For brevity, we only refer to 'significant' and 'not significant' in this table, while in the main text more attention is paid to effect sizes. Since we use two-sided tests, we can interpret effects in the opposite direction from that predicted. To estimate whether an effect will be small enough to be considered negligible, we conduct equivalence tests using the two one-sided tests (TOST) method[90,91], implemented by computing the 90% CI around the estimate of the regression coefficient. We consider an effect to be negligible if it is lower than an effect corresponding to a regression coefficient of 0.1, computed after having standardized every variable. This corresponds to a Cohen's d of 0.2 between the two conditions, which we consider to be the minimal effect size of interest. The scaling is done separately for each comparison. For instance, the scaling for the comparison between the control condition and consensus condition will be done based only on the participants of these two conditions. We do so to make the meaning of the regression coefficient as similar as possible to the meaning of Cohen's d. If we find a significant difference in the expected direction, we will conclude that being exposed to the consensus led to more positive attitudes towards GMOs than reading a description of what a GMO is, and that H1 is supported. If we find a significant difference in the direction opposite to that predicted, we will conclude that H1 is not supported and that the opposite of H1 is supported (in this case, that reading a GMO description led to more attitude change than being exposed to the consensus). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% confidence Interval around the regression coefficient of condition; if this confidence interval includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.
H2	Participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the control condition, controlling for their initial attitudes towards GMOs.			If we find a significant difference in the expected direction, we will conclude that reading the counterarguments and rebuttals available on the chatbot led to more positive attitudes towards GMOs than reading a description of what a GMO is, and that H2 is supported. If we find a significant difference in the direction opposite to that predicted, we will conclude that H2 is not supported and that the opposite of H2 is supported (in this case, that reading counterarguments and rebuttals available on the chatbot led to more attitude change than reading a description of what a GMO is). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% confidence interval around the regression coefficient of condition; if this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.

Continued

Table 1 | Design table (continued)

	Hypothesis	Sampling (for example, power analysis)	Analysis	Interpretation given to different outcomes
H3	Participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the control condition, controlling for their initial attitudes towards GMOs.			If we find a significant difference in the expected direction, we will conclude that interacting with the chatbot led to more positive attitudes towards GMOs than reading a description of what a GMO is, and that H3 is supported. If we find a significant difference in the direction opposite to that predicted, we will conclude that H3 is not supported and that the opposite of H3 is supported (in this case, that interacting with the chatbot led to more positive attitudes towards GMOs change than reading a description of what a GMO is). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% CI around the regression coefficient of condition; if this confidence interval includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.
H4	Participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the consensus condition, controlling for their initial attitudes towards GMOs.		To test H4, based on the same regression model as in H1–3, we conduct a linear contrast analysis between the counterarguments condition and the consensus condition. H4 leads us to expect that the consensus condition will predict less attitude change in the direction of more positive attitudes towards GMOs than the counterarguments condition. To test H5 based on the same regression model as in H1–3, we conduct a linear contrast analysis between the chatbot and the counterarguments conditions. H5 leads us to expect that the counterarguments condition will predict less attitude change in the direction of more positive attitudes towards GMOs than the chatbot condition.	If we find a significant difference in the expected direction, we will conclude that reading the counterarguments and rebuttals available on the chatbot led to more positive attitudes towards GMOs than being exposed to the scientific consensus, and that H4 is supported. If we find a significant difference in the direction opposite to that predicted, we will conclude that H4 is not supported and that the opposite of H4 is supported (in this case, that being exposed to the scientific consensus on GMOs led to more positive attitudes towards GMOs change than reading the counterarguments and rebuttals available on the chatbot). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% CI around the regression coefficient of condition. If this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.
H5	Participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the counterarguments condition, controlling for their initial attitudes towards GMOs.			If we find a significant difference in the expected direction, we will conclude that interacting with the chatbot led to more positive attitudes towards GMOs than reading the counterarguments and rebuttals available on the chatbot without being able to interact with it, and that H5 is supported. If we find a significant difference in the direction opposite to that predicted, we will conclude that H5 is not supported and that the opposite of H5 is supported (in this case, that interacting with the chatbot led to less positive attitude change toward GMOs than reading the arguments available on the chatbot). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% CI around the regression coefficient of condition. If this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.

Continued

Table 1 | Design table (continued)

	Hypothesis	Sampling (for example, power analysis)	Analysis	Interpretation given to different outcomes
H6	In the chatbot condition, the number of arguments explored by participants will predict holding more positive attitudes towards GMOs after the experimental task, controlling for their initial attitudes towards GMOs (that is, exploring more arguments should lead to more positive attitude change).		To test H6, we conduct one multivariate regression among participants in the chatbot condition, with attitudes after the experimental task as the dependent variable and attitudes before the experimental task together with the total number of arguments explored by participants as predictors.	If the number of arguments explored by participants significantly predicts more positive attitudes toward GMOs, we will conclude that, the more arguments participants are exposed to, the more they change their minds in favour of GMOs, supporting H6. If we find a significant positive difference in the direction opposite to that predicted, we will conclude that H6 is not supported and that the opposite of H6 is supported (in this case, that being exposed to fewer arguments led to more positive attitude change toward GMOs). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% CI around the regression coefficient of number of arguments. If this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.
H7	In the chatbot condition, time spent on the task should lead to more positive attitudes towards GMOs after the experimental task than time spent on the counterarguments condition, controlling for initial attitudes towards GMOs.		To test H7, we conduct one multivariate regression among participants in the chatbot condition and the counterarguments condition, with attitudes after the experimental task as the dependent variable, and using the time variable, the condition variable and an interaction between the time variable and the condition variable as predictors. The time variable is mean-centred to facilitate interpretation of the regression coefficients.	If time spent interacting with the chatbot led to more attitude change in favour of GMOs in the chatbot condition than in the counterarguments condition, we will conclude that being exposed to only relevant counterarguments is more efficient at changing people's mind in favour of GMOs than presenting them with potentially irrelevant arguments. If we find a significant difference in the direction opposite to that predicted, we will conclude that H7 is not supported and that the opposite of H7 is supported (in this case, that not interacting with the chatbot was more efficient at changing people's mind in favour of GMOs). If we find no significant effect, we will conclude that we did not find support for the hypothesis. If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we will compute the 90% CI around the regression coefficient of the interaction between time and condition; if this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.
H8	In the chatbot condition, participants will hold more positive attitudes after the experimental task on issues for which they have explored more of the rebuttals related to the issue, controlling for their initial attitudes on the issues, the type of issue and the total (that is, related to the issue or not) number of arguments explored.		To test H8, we count the number of arguments explored in each of the four branches (between zero and nine). To investigate the relation between the type of arguments that participants explored on the chatbot and attitude change on these particular aspects of GMOs (health, ecology, economy and usefulness), we conduct a linear mixed-effects model with participants as random effect (varying intercepts), attitudes after the experimental task on a specific issue as the dependent variable and number of arguments explored on the same specific issue, together with attitudes before the experimental task on the same issue, the total number of arguments explored and the type of issue as predictors.	If, after having interacted with the chatbot, participants hold more positive attitudes on issues for which they have explored more of the rebuttals related to the issue, we will conclude that participants paid attention to the content of the arguments and probably changed their mind because of arguments' content and did not use a low-level heuristic in which they are convinced by the sheer number of arguments. If we find a significant difference in the direction opposite to that predicted, we will conclude that H8 is not supported and that the opposite of H8 is supported. If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, we compute the 90% CI around the regression coefficient of the issues explored by participants; if this confidence interval includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible.

Continued

Table 1 | Design table (continued)

	Hypothesis	Sampling (for example, power analysis)	Analysis	Interpretation given to different outcomes
H9	H1, H2 and H3 also hold true among the third of the participants initially holding the most negative attitudes about GMOs. (Note that this criterion is more stringent than an absence of backfire effect, as it claims that there will be a positive effect even among participants with the most negative initial attitudes.)		To test H9, we conduct the same analysis used to test H1, H2 and H3 (that is, a multivariate regression with attitudes after the experimental as dependent variable and attitudes before the experimental task together with condition as predictors, with the control condition as the baseline for the variable condition) among the one-third of participants initially holding the most negative attitudes toward GMOs. If H1, H2 and H3 hold true among the third of the participants initially holding the most negative attitudes about GMOs, we will conclude that we did not find evidence in favour of the backfire effect and that attitude change was not dissimilar between the third of the participants initially holding the most negative attitudes about GMOs and the rest of participants. If participants holding the most negative attitudes about GMOs showed more positive attitude change toward GMOs than the rest of participants H9 will still be supported. On the other hand, if the effect goes in the direction opposite to that predicted, we will conclude that the opposite of H9 is supported (that is, it will be evidence in favour of a backfire effect). If we find no significant difference, we will conclude that we cannot reject the null hypothesis. Then, using the TOST method, will compute the 90% CI around the regression coefficient of condition for each hypothesis (that is, H1–3). If this CI includes neither the value of −0.1 nor the value of 0.1, we will declare the effect to be practically negligible. Again, due to the reduced power and the lack of correction to test H9, we will be especially cautious in interpreting the results of H9. Except for H9, all the hypotheses presuppose that participants will, on average, either not change their opinion between the pre- and the post-treatment questions or that they will become more favourable towards GMOs (that is, there is either a significantly positive change or a significant absence of difference, as tested with an equivalence test). If, on the contrary, we find that, in one or more condition, participants became significantly less favourable towards GMOs, this would provide evidence for a backfire effect.

a large effect of the chatbot on attitude change (ANOVA, $\eta^2 = 0.15$). However, because the current pre-registered study compared the chatbot with controls, where some attitude change occurred, we expected the effect to be smaller (small to medium instead of large).

We performed an a priori power analysis with G*Power3[75]. To compute the necessary number of participants, we decided that the minimal effect size of interest would correspond to a Cohen's d of 0.2 between two different experimental conditions, since this corresponds to what is generally seen as a small effect[76]. Based on a correlation of 0.75 between the initial and final GMOs attitudes (estimated from the pilot), we needed at least 275 participants per condition to detect this effect, at an α level of 5% and a power of 95%, and based on a two-tailed test (see Power analysis section in Supplementary information for more details). We expected that approximatively 15% of participants would encounter problems accessing the chatbot's interface (a percentage estimated while pre-testing the chatbot). We planned to exclude these participants. To anticipate the losses in participants unable to access the chatbot, we planned on recruiting 324 participants (275/0.85) instead of 275 in the chatbot condition and in the counterarguments condition. We planned to recruit a total of 1,198 UK participants on the crowd-sourcing platform Prolific Academic. Data collection stopped when each condition reached the minimum number of participants required by the power analysis after exclusions (due to inability to access the chatbot).

During the stage 2 peer-review process, reviewers noticed that our a priori power analysis at stage 1 had not taken into account our correction for multiple testing and that it was based on a comparison of two conditions instead of four. To enable readers to evaluate the impact of this on our power calculations, we repeated our simulations with fixed sample sizes and a fixed effect size (the a priori defined minimum effect size of interest, that is, $d = 0.2$). These simulations showed that the actual power to detect the smallest effect size of theoretical interest was higher than 91% in all conditions (see Supplementary information for further details).

Participants. Between 15 October 2020 and 26 October 2020, we recruited 1,306 participants (paid £1.38) from the United Kingdom on Prolific Academic. We excluded 156 participants who could not, or did not, access the chatbot, leaving 1,150 participants in total (776 women, $M_{age} = 34.74$, s.d. 12.87) with 302 participants in the chatbot condition, 299 participants in the counterarguments condition, 273 participants in the consensus condition and 275 participants in the control condition.

Analysis plan. All analyses were conducted with R (version 3.6.1)[87], using R Studio (version 1.2.5019)[88]. All statistical tests are two-sided. We refer to 'statistically significant' as P value being lower than an α of 0.05. We controlled for multiple comparisons by applying the Benjamini–Hochberg method to H1–8 (which controls for the false discovery rate and has a less negative impact on statistical

power than alternative methods), but not to H9, for two reasons. First, we had planned on testing H9 only if one of the first three hypotheses were supported. As a consequence, H9 does not increase the familywise error rate (this is a special case of the closure principle in multiple comparisons[89]). Second, since H9 was conducted only on a third of the participants, controlling for multiple comparisons would have reduced our statistical power even more. Due to this reduced power and the lack of correction, we were especially cautious in interpreting the results of H9.

All the P values reported in the exploratory analyses have been corrected for multiple comparisons by applying the Benjamini–Hochberg method. This correction included the P values of the confirmatory analyses (whereas the correction applied to the P values of the confirmatory analyses did not include the P values of the exploratory analyses). We used this method to maximize power for the confirmatory analyses (and conform to the pre-registered plan) while limiting the risk of false positives for the exploratory analyses.

Given that we used null hypothesis statistical testing, null results were interpreted as the impossibility to reject H0, and as an absence of support for the hypothesis tested, but not as support for H0. Data from previous studies suggested that our experimental design would allow us to test our hypotheses. First, survey data on attitudes about GMOs in the UK, or other European countries, suggested that participants would be far from the ceiling (that is, being maximally in favour of GMOs)[40,77], so that we would be able to observe attitude change towards attitudes more favourable to GMOs. Second, previous studies using consensus messaging[9,10] suggested that some attitude change should be observed in our consensus condition, which could thus be used as a positive control.

We compared participants' attitudes before and after being exposed to one of the four conditions by using a composite measure composed of the mean ratings of the four GMOs attitudes questions. To make our measures more intuitive, we reverse-coded all but one of the questions (the first) such that higher numbers denote a more positive attitude towards GMOs. Time was measured by our custom-made website that provides a precise and reliable measure of the time spent by participants interacting with the chatbot in the chatbot condition or reading the arguments in the counterarguments condition. To estimate whether an effect was small enough to be considered negligible, we conducted equivalence tests using the two one-sided tests (TOST) method[90,91], which we implemented by computing the 90% CI around the estimate of the regression coefficient. To further assist with interpretation, we provided Bayes factors at stage 2 for all confirmatory analyses (BFs were not registered at stage 1). To calculate the Bayes factors, we used the 'lmBF' function from the BayesFactor R package (package 'bayesfactor'[92]). The choice of priors and Markov chain Monte Carlo settings were the default settings of the package (as of 2020). The R script used to analyse the data, together with the mock dataset on which the script was tested, are available at https://osf.io/cb7wf/.

According to H1, participants will hold more positive attitudes towards GMOs after the experimental task in the consensus condition than in the control condition, when controlling for their initial attitudes towards GMOs.

According to H2, participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the control condition, when controlling for their initial attitudes towards GMOs.

According to H3, participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the control condition, when controlling for their initial attitudes towards GMOs.

H1–3 were tested on the full dataset with one multivariate regression. Attitudes after the experimental task were set as the dependent variable, while attitudes before the experimental task and condition were set as predictors. The control condition was set as the baseline for the variable condition. In other words, consensus condition, counterarguments condition and chatbot condition were each compared with the control condition. Attitudes before the experimental task was mean-centred to facilitate the interpretation of the intercept, which corresponds to the mean post-attitude for the control condition.

According to H4, participants will hold more positive attitudes towards GMOs after the experimental task in the counterarguments condition than in the consensus condition, controlling for their initial attitudes towards GMOs.

H4 was based on the same regression model as in H1–3. We conducted a linear contrast analysis between the counterarguments condition and the consensus condition. H4 led us to expect that the consensus condition would predict less attitude change in the direction of more positive attitudes towards GMOs than the counterarguments condition.

According to H5, participants will hold more positive attitudes towards GMOs after the experimental task in the chatbot condition than in the counterarguments condition, controlling for their initial attitudes towards GMOs

H5 was based on the same regression model as in H1–3. We conducted a linear contrast analysis between the chatbot condition and the counterarguments condition. H5 led us to expect that the counterarguments condition would predict less attitude change in the direction of more positive attitudes towards GMOs than the chatbot condition.

According to H6, in the chatbot condition, the number of arguments explored by the participants will predict holding positive attitudes towards GMOs after the experimental task when controlling for their initial attitudes towards GMOs (that is, exploring more arguments should lead to more positive attitude change).

H6 was tested with one multivariate regression among participants in the chatbot condition, with attitudes after the experimental task as the dependent variable, and attitudes before the experimental task together with the total number of arguments explored by participants as predictors.

According to H7, in the chatbot condition, time spent on the task should lead to more positive attitude towards GMOs after the experimental task than time spent on the counterarguments condition, when controlling for their initial attitudes towards GMOs.

H7 was tested with one multivariate regression among participants in the chatbot condition and the counterarguments condition, with attitudes after the experimental task as the dependent variable, and using the time variable, the condition variable and an interaction between the time variable and the condition variable as predictors. The time variable was mean-centred to facilitate the interpretation of the regression coefficients.

The four questions measuring attitudes towards GMOs correspond to concerns about GM foods safety, GMOs' ecological impact, GMO's usefulness and the socio-economic dimension of GMOs. The internal consistency of the scale was higher after the treatment ($\alpha = 0.79$) than before the treatment ($\alpha = 0.68$), but this effect was mostly driven by the chatbot (pre, 0.67; post, 0.79) and counterarguments condition (pre, 0.66; post, 0.81) rather than the control (pre, 0.70; post, 0.71) and consensus condition (pre, 0.68; post, 0.71). The chatbot menu is also composed of four main counterarguments against GMOs: 'GMOs might not be safe to eat', which targets health concerns, 'GMOs could hurt the planet', which targets ecological concerns, 'The way GMOs are commercialized is problematic', which targets economic concerns, and 'We don't really need GMOs', which targets the usefulness of GMOs. Each of these main counterarguments is answered by a rebuttal, which can then be answered by several counterarguments, which have their own rebuttals, and so forth. According to H8, on the chatbot condition, participants will hold more positive attitudes after the experimental task on issues for which they have explored more of the relevant rebuttals targeted at the issue, when controlling for their initial attitudes on the issues and the total number of arguments explored (not necessarily related to the issue). To test H8, we counted the number of arguments explored in each of the four branches (between zero and nine).

To investigate the relation between the type of arguments that participants explored on the chatbot and attitude change on these particular aspects of GMOs (health, ecology, economy and usefulness), we conducted a linear mixed-effects model with participants as random effect (varying intercepts), attitudes after the experimental task on a specific issue as the dependent variable and number of arguments explored on the same specific issue, together with attitudes before the experimental task on the same issue, the total number of arguments explored and the type of issue as predictors.

According to H8, we expect that the number of arguments explored by participants on the topic related to the subscale will be a predictor of more favourable attitudes towards GMOs after having interacted with the chatbot, when controlling for their initial attitudes towards GMOs.

According to H9, the hypotheses H1, H2 and H3 should hold true among the third of the participants most opposed to GMOs.

H9 was tested by conducting the same analysis used to test H1, H2 and H3 (that is, a multivariate regression with attitudes after the experimental as dependent variable, and attitudes before the experimental task together with condition as predictors, with the control condition as the baseline for the variable condition) among the one-third of participants initially holding the most negative attitudes toward GMOs.

We made no predictions regarding gender, education or other socio-demographic variables. We did not add these variables in the models since their influence should mostly be taken into account when controlling for initial attitudes.

The sentence 'Although it is methodologically impossible to completely disentangle the effects of the mode of presentation (for example degree of interactivity) and of the specific information presented, the present experiment provides the first test of whether addressing people's counterarguments, in particular by using an interactive chatbot, results in attitude changes that are larger than those obtained with a common messaging technique.' was modified in the stage 1 portion of the manuscript before acceptance at stage 2 on editorial request, in order to comply with journal policies regarding priority claims.

The following paragraph 'However these studies remain limited in particular as they did not include control groups comparable to the present control conditions. Instead the chatbots were compared either (i) to argumentation between participants[31] (ii) to a chatbot that could not address all counterarguments[30] or (iii) to a chatbot that did not take into account users' counterarguments[32]. Moreover the robustness of these results is questionable since their design had poor sensitivity to detect even large effect sizes of dz = 0.5 (the study with the greatest sensitivity[32], which recruited 25 participants per condition on average, had no more than 0.71 power to detect large effects (dz = 0.5) with an alpha of 0.05; in the other studies power was even lower: 0.49[31] and 0.52[30]).' was removed from the stage 2 submission because, since in principle acceptance of the stage 1 protocol, studies that are not subject to these limitations have been published[69].

Data availability

Code availability

References

1. *Public and Scientists' Views on Science and Society* (Pew Research Center, 2015).
2. Nyhan, B. & Reifler, J. When corrections fail: the persistence of political misperceptions. *Polit. Behav.* **32**, 303–330 (2010).
3. Nyhan, B., Reifler, J., Richey, S. & Freed, G. L. Effective messages in vaccine promotion: a randomized trial. *Pediatrics* **133**, e835–e842 (2014).
4. Cook, J. & Lewandowsky, S. *The Debunking Handbook* (Sevloid Art, 2011).
5. Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N. & Cook, J. Misinformation and its correction: continued influence and successful debiasing. *Psychol. Sci. Public Interest* **13**, 106–131 (2012).
6. Guess, A. & Coppock, A. Does counter-attitudinal information cause backlash? Results from three large survey experiments. *Br. J. Polit. Sci.* https://doi.org/10.1017/S0007123418000327 (2018).
7. Wood, T. & Porter, E. The elusive backfire effect: mass attitudes' steadfast factual adherence. *Polit. Behav.* **41**, 135–163 (2019).
8. Landrum, A. R., Hallman, W. K. & Jamieson, K. H. Examining the impact of expert voices: communicating the scientific consensus on genetically-modified organisms. *Environ. Commun.* https://doi.org/10.1080/17524032.2018.1502201 (2018).
9. Dixon. Applying the gateway belief model to genetically modified food perceptions: new insights and additional questions. *J. Commun.* https://doi.org/10.1111/jcom.12260 (2016).

10. Kerr, J. R. & Wilson, M. S. Changes in perceived scientific consensus shift beliefs about climate change and GM food safety. *PLoS ONE* **13**, e0200295 (2018).

11. Claidière, N., Trouche, E. & Mercier, H. Argumentation and the diffusion of counter-intuitive beliefs. *J. Exp. Psychol. Gen.* **146**, 1052–1066 (2017).

12. Trouche, E., Sander, E. & Mercier, H. Arguments, more than confidence, explain the good performance of reasoning groups. *J. Exp. Psychol. Gen.* **143**, 1958–1971 (2014).

13. Laughlin, P. R. *Group Problem Solving* (Princeton Univ. Press, 2011).

14. Minson, J. A., Liberman, V. & Ross, L. Two to tango. *Personal. Soc. Psychol. Bull.* **37**, 1325–1338 (2011).

15. Smith, M. K. et al. Why peer discussion improves student performance on in-class concept questions. *Science* **323**, 122–124 (2009).

16. Mercier, H. The argumentative theory: predictions and empirical evidence. *Trends Cogn. Sci.* **20**, 689–700 (2016).

17. Mercier, H. & Sperber, D. *The Enigma of Reason* (Harvard Univ. Press, 2017).

18. Minozzi, W., Neblo, M. A., Esterling, K. M. & Lazer, D. M. Field experiment evidence of substantive, attributional, and behavioral persuasion by members of Congress in online town halls. *Proc. Natl Acad. Sci. USA* **112**, 3937–3942 (2015).

19. Broockman, D. & Kalla, J. Durably reducing transphobia: a field experiment on door-to-door canvassing. *Science* **352**, 220–224 (2016).

20. Altay, S. & Lakhlifi, C. Are science festivals a good place to discuss heated topics? *J. Sci. Commun.* **19**, A07 (2020).

21. Chanel, O., Luchini, S., Massoni, S. & Vergnaud, J.-C. Impact of information on intentions to vaccinate in a potential epidemic: swine-origin influenza A (H1N1). *Soc. Sci. Med.* **72**, 142–148 (2011).

22. Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H. & Holowchak, M. Reasoning in conversation. *Cogn. Instr.* **11**, 347–364 (1993).

23. Edwards, K. & Smith, E. E. A disconfirmation bias in the evaluation of arguments. *J. Personal. Soc. Psychol.* **71**, 5–24 (1996).

24. Greenwald, A. G. in *Psychological Foundations of Attitudes* (eds. Greenwald, A. G., Brock, T. C. & Ostrom, T. M.) 147–170 (Academic Press, 1968).

25. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Polit. Sci.* **50**, 755–769 (2006).

26. Trouche, E., Shao, J. & Mercier, H. Objective evaluation of demonstrative arguments. *Argumentation* **33**, 23–43 (2019).

27. Mercier, H., Bonnier, P. & Trouche, E. in *Cognitive Unconscious and Human Rationality* (eds. Macchi, L., Bagassi, M. & Viale, R.) 205–218 (MIT Press, 2016).

28. Fay, N., Garrod, S. & Carletta, J. Group discussion as interactive dialogue or as serial monologue: the influence of group size. *Psychol. Sci.* **11**, 481–486 (2000).

29. Krems, J. A. & Wilkes, J. Why are conversations limited to about four people? A theoretical exploration of the conversation size constraint. *Evol. Hum. Behav.* **40**, 140–147 (2019).

30. Andrews, P., Manandhar, S. & De Boni, M. in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue* 138–147 (2008). https://doi.org/10.3115/1622064.1622093

31. Rosenfeld, A. & Kraus, S. in *Proceedings of the Twenty-Second European Conference on Artificial Intelligence* 320–328 (IOS Press, 2016). https://doi.org/10.3233/978-1-61499-672-9-320

32. Chalaguine, L. A., Hunter, A., Hamilton, F. L. & Potts, H. W. Impact of argument type and concerns in argumentation with a chatbot. in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* https://doi.org/10.1109/ICTAI.2019.00224 (2019).

33. Baulcombe, D., Dunwell, J., Jones, J., Pickett, J. & Puigdomenech, P. *GM Science Update: a Report to the Council for Science and Technology* (2014). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/292174/cst-14-634a-gm-science-update.pdf

34. *A Decade of EU-Funded GMO Research* (European Commission, 2010).

35. National Academies of Sciences and Medicine. *Genetically Engineered Crops: Experiences and Prospects* (National Academies Press, 2016).

36. Nicolia, A., Manzo, A., Veronesi, F. & Rosellini, D. An overview of the last 10 years of genetically engineered crop safety research. *Crit. Rev. Biotechnol.* **34**, 77–88 (2014).

37. Ronald, P. Plant genetics, sustainable agriculture and global food security. *Genetics* **188**, 11–20 (2011).

38. *Statement by the AAAS Board of Directors on Labeling of Genetically Modified Foods* (American Assoiation for the Advancement of Science, 2012).

39. Yang, Y. T. & Chen, B. Governing GMOs in the USA: science, law and public health. *J. Sci. Food Agric.* **96**, 1851–1855 (2016).

40. Bonny, S. Why are most Europeans opposed to GMOs?: factors explaining rejection in France and Europe. *Electron. J. Biotechnol.* **6**, 7–8 (2003).

41. Cui, K. & Shoemaker, S. P. Public perception of genetically-modified (GM) food: a nationwide Chinese consumer study. *npj Sci. Food* **2**, 10 (2018).

42. Gaskell, G., Bauer, M. W., Durant, J. & Allum, N. C. Worlds apart? The reception of genetically modified foods in Europe and the US. *Science* **285**, 384–387 (1999).

43. Scott, S. E., Inbar, Y. & Rozin, P. Evidence for absolute moral opposition to genetically modified food in the United States. *Perspect. Psychol. Sci.* **11**, 315–324 (2016).

44. *Baromètre sur la Perception des Risques et de la Sécurité par les Français* (IRSN, 2017).

45. *Les Français et les OGM* (Ifop, 2012).

46. Burke, D. GM food and crops: what went wrong in the UK? *EMBO Rep.* **5**, 432–436 (2004).

47. Poortinga, W. & Pidgeon, N. *Public Perceptions of Genetically Modified Food and Crops, and the GM Nation? Public Debate on the Commercialisation of Agricultural Biotechnology in the UK: Main Findings of a British Survey* (Centre for Environmental Risk, 2004).

48. Cordon, G. GM crops opposition may have been 'over-estimated'. *The Scotsman* (19 February 2004).

49. Bonny, S. Will Biotechnology Lead to More Sustainable Agriculture? in *Proc. of NE-165 Conference* (2000).

50. Hielscher, S., Pies, I., Valentinov, V. & Chatalova, L. Rationalizing the GMO debate: the ordonomic approach to addressing agricultural myths. *Int. J. Environ. Res. Public Health* **13**, 476 (2016).

51. Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C. & Leiserowitz, A. Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nat. Clim. Change* **1**, 462 (2011).

52. Dunwoody, S. & Kohl, P. A. Using weight-of-experts messaging to communicate accurately about contested science. *Sci. Commun.* **39**, 338–357 (2017).

53. Kohl, P. A. et al. The influence of weight-of-evidence strategies on audience perceptions of (un)certainty when media cover contested science. *Public Understand. Sci.* **25**, 976–991 (2016).

54. Lewandowsky, S., Gignac, G. E. & Vaughan, S. The pivotal role of perceived scientific consensus in acceptance of science. *Nat. Clim. Change* **3**, 399–404 (2013).

55. van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D. & Maibach, E. W. The scientific consensus on climate change as a gateway belief: experimental evidence. *PLoS ONE* **10**, e0118489 (2015).

56. van der Linden, S. L., Leiserowitz, A. & Maibach, E. Gateway illusion or cultural cognition confusion? *J. Sci. Commun.* https://doi.org/10.22323/2.16050204 (2017).

57. Petty, R. E. & Cacioppo, J. T. in *Advances in Experimental Social Psychology* (ed. Berkowitz, L.) 123–205 (Academic Press., 1986).

58. Petty, R. E. & Cacioppo, J. T. The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion. *J. Personal. Soc. Psychol.* **46**, 69 (1984).

59. Ecker, U. K. H. & Ang, L. C. Political attitudes and the processing of misinformation corrections. *Polit. Psychol.* **40**, 241–260 (2019).

60. Kahan, D. Ideology, motivated reasoning, and cognitive reflection. *Judgm. Decis. Mak.* **8**, 407–424 (2013).

61. Kahan, D., Jenkins-Smith, H. & Braman, D. Cultural cognition of scientific consensus. *J. Risk Res.* **14**, 147–174 (2011).

62. Schmid, P. & Betsch, C. Effective strategies for rebutting science denialism in public discussions. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-019-0632-40 (2019).

63. van der Linden, S., Leiserowitz, A. & Maibach, E. The gateway belief model: a large-scale replication. *J. Environ. Psychol.* **62**, 49–58 (2019).

64. van der Linden, S., Maibach, E. & Leiserowitz, A. Exposure to scientific consensus does not cause psychological reactance. *Environ. Commun.* https://doi.org/10.1080/17524032.2019.1617763 (2019).

65. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 5 (2014).

66. Bullock, J. G., Green, D. P. & Ha, S. E. Yes, but what's the mechanism? (don't expect an easy answer). *J. Personal. Soc. Psychol.* **98**, 550–558 (2010).

67. McPhetres, J., Rutjens, B. T., Weinstein, N. & Brisson, J. A. Modifying attitudes about modified foods: increased knowledge leads to more positive attitudes. *J. Environ. Psychol* https://doi.org/10.1016/j.jenvp.2019.04.012 (2019).

68. Hasell, A., Lyons, B. A., Tallapragada, M. & Jamieson, K. H. Improving GM consensus acceptance through reduced reactance and climate change-based message targeting. *Environ. Commun.* https://doi.org/10.1080/17524032.2020.1746377 (2020).

69. Altay, S., Hacquin, A.-S., Chevallier, C. & Mercier, H. Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. *J. Exp. Psychol. Appl.* https://doi.org/10.1037/xap0000400 (2021).

70. Swire-Thompson, B., DeGutis, J. & Lazer, D. Searching for the backfire effect: measurement and design considerations. https://doi.org/10.1016/j.jarmac.2020.06.006 (2020).

71. Bode, L., Vraga, E. K. & Tully, M. Correcting misperceptions about genetically modified food on social media: examining the impact of experts, social media heuristics, and the gateway belief model. *Sci. Commun.* **43**, 225–251 (2021).

72. Bode, L. & Vraga, E. K. In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *J. Commun.* **65**, 619–638 (2015).

73. Vraga, E. K. & Bode, L. Using expert sources to correct health misinformation in social media. *Sci. Commun.* **39**, 621–645 (2017).

74. Coppock, A., Leeper, T. J. & Mullinix, K. J. Generalizability of heterogeneous treatment effect estimates across samples. *Proc. Natl Acad. Sci. USA* **115**, 12441–12446 (2018).

75. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res Methods* **39**, 175–191 (2007).

76. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn (Lawrence Erlbaum Assoicates, 1988).

77. Bonny, S. Factors explaining opposition to GMOs in France and the rest of Europe. *Consumer Acceptance of Genetically Modified Foods* 169 (2004).

78. Evenson, R. E. & Santaniello, V. *Consumer Acceptance of Genetically Modified Foods* (CABI, 2004).

79. McHughen, A. & Wager, R. Popular misconceptions: agricultural biotechnology. *N. Biotechnol.* **27**, 724–728 (2010).

80. Parrott, W. Genetically modified myths and realities. *N. Biotechnol.* **27**, 545–551 (2010).

81. Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J. & Van Montagu, M. Fatal attraction: the intuitive appeal of GMO opposition. *Trends Plant Sci.* https://doi.org/10.1016/j.tplants.2015.03.011 (2015).

82. Key, S., Ma, J. K. & Drake, P. M. Genetically modified plants and human health. *J. R. Soc. Med.* **101**, 290–298 (2008).

83. Klümper, W. & Qaim, M. A meta-analysis of the impacts of genetically modified crops. *PLoS ONE* **9**, e111629 (2014).

84. Pellegrino, E., Bedini, S., Nuti, M. & Ercoli, L. Impact of genetically engineered maize on agronomic, environmental and toxicological traits: a meta-analysis of 21 years of field data. *Sci. Rep.* **8**, 3113 (2018).

85. Snell, C. et al. Assessment of the health impact of GM plant diets in long-term and multigenerational animal feeding trials: a literature review. *Food Chem. Toxicol.* **50**, 1134–1148 (2012).

86. Brysbaert, M. How many words do we read per minute? A review and meta-analysis of reading rate. *J. Mem. Lang.* https://doi.org/10.1016/j.jml.2019.104047 (2019).

87. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).

88. *RStudio: Integrated Development for R* (RStudio, 2015).

89. Bretz, F., Hothorn, T. & Westfall, P. *Multiple Comparisons Using R* (CRC Press, 2016).

90. Lakens, D. Equivalence tests: a practical primer for *t* tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).

91. Campbell, H. Equivalence testing for standardized effect sizes in linear regression. *arXiv* https://arxiv.org/abs/2004.01757 (2020).

92. Morey, R. D., Rouder, J. N., Jamil, T. & Morey, M. R. D. Package 'bayesfactor'. http://www.cran.r-projectorg/web/packages/BayesFactor/BayesFactorpdfi (2015).

Author contributions

S.A., M.S., A.-S.H. and H.M. conceived and designed the experiments. S.A., A.-S.H. and H.M. performed the experiments. S.A., A.-S.H., A.A. and H.M. analysed the data. S.A., M.S., A-S.H., A.A., S.B. and H.M. contributed materials/analysis tools. S.A., A.-S.H., A.A. and H.M. wrote the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01271-w.

Correspondence and requests for materials should be addressed to Sacha Altay or Hugo Mercier.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

Corresponding author(s):   Sacha Altay & Hugo Mercier

Last updated by author(s):   Nov 3, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We collected data with Qualtrics and recruited participants via Prolific Academic.
Data analysis	All analyses were conducted with R (v.3.6.1), using R Studio (v.1.2.5019). The R scripts associated with this research are available on OSF at: https://osf.io/cb7wf/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data associated with this research, together with the code of the chatbot and the materials, are available on OSF at: https://osf.io/cb7wf/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative & Experimental.
Research sample	We recruited U.K. participants, not representative of the general population. 776 women out of 1150participants. Age, M = 34.74, SD = 12.87. We were not interested in testing the effect of particular demographic variables. Generally convenience samples offer similar results to representative samples in terms of experimental manipulations.
Sampling strategy	We performed a power analysis, see: 'We performed an a priori power analysis with G*Power3 75. To compute the necessary number of participants, we decided that the minimal effect size of interest would correspond to a Cohen's d of 0.2 between two different experimental conditions, since this corresponds to what is generally seen as a small effect 76. Based on a correlation of 0.75 between the initial and final GMO attitudes (estimated from the pilot), we needed at least 275 participants per condition to detect this effect, at an α-level of 5%, a power of 95%, and based on a two-tailed test (see SI for more details). We expected that approximatively 15% of participants would encounter problems accessing the chatbot's interface (a percentage estimated while pre-testing the chatbot). We planned to exclude these participants. To anticipate the losses in participants unable to access the chatbot, we planned on recruiting 324 participants (275/0.85) instead of 275 in the Chatbot Condition and in the Counterarguments Condition. We planned to recruit a total of 1198 UK participants on the crowdsourcing platform Prolific Academic. Data collection stopped when each condition reached the minimum number of participants required by the power analysis after exclusions (due to inability to access the chatbot).'
Data collection	'Since data collection was automatized on Qualtrics and Prolific Academic, that all our statistical analyses were pre-registered, and that there was no subjective coding of the data, the experimenters were not blind to the conditions of the experiments. Participants were not blinded to the study hypotheses.'
Timing	'Between the 15th of October 2020 and the 26th of October 2020, we recruited 1306 participants (paid £1.38) from the United Kingdom on Prolific Academic.
Data exclusions	'We excluded 156 participants who could not, or did not, access the chatbot. Leaving 1150 participants in total (776 women, MAge = 34.74, SD = 12.87)—302 participants in the Chatbot Condition, 299 participants in the Counterarguments Condition, 273 participants in the Consensus Condition, and 275 participants in the Control Condition.'
Non-participation	37 participants started the experiments but did not finish it.
Randomization	'Participants were randomly assigned to one of the four conditions by a pseudo-randomizer on the survey platform "Qualtrics" (i.e. a randomizer that ensures that an equal number of participants is attributed to each condition).'

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Population characteristics	See above.
Recruitment	Participants were recruited through the platform Prolific Academic.
Ethics oversight	CER-Paris Descartes; N° 2019-03-MERCIER

Note that full information on the approval of the study protocol must also be provided in the manuscript.