Argument evaluation and production in the correction of political innumeracy
Martin Dockendorff, Hugo Mercier

To cite this version:
Martin Dockendorff, Hugo Mercier. Argument evaluation and production in the correction of political innumeracy. Thinking and Reasoning, In press. hal-04149831

HAL Id: hal-04149831
https://cnrs.hal.science/hal-04149831
Submitted on 10 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright
Argument evaluation and production in the correction of political innumeracy

Martin Dockendorff\textsuperscript{1} & Hugo Mercier\textsuperscript{2}* 

\textsuperscript{1}Department of Cognitive Science, Central European University, Vienna, Austria.

\textsuperscript{2}Institut Jean Nicod, Département d'Études Cognitives, ENS, EHESS, PSL University, CNRS, Paris, France.

* To whom correspondence should be addressed:

Hugo Mercier Institut Jean Nicod ENS 29 rue d’Ulm, 75005 Paris France
hugo.mercier@gmail.com

Accepted at \textit{Thinking & Reasoning}

Non-proofread version, please do not cite
Abstract.

The public is largely innumerate, making systematic mistakes in estimating some politically relevant facts, such as the share of foreign-born citizens. In two-step or multistep flow models, such mistakes could be corrected if better-informed citizens were able to convince their peers, in particular by using good arguments citing reliable sources. In six experiments, we find two issues that dampen the potential power of this two-step flow process. First, even though participants were more convinced by good than by poor arguments, many did not change their minds, even when confronted with good arguments. Second, participants are not inclined to spontaneously generate arguments that cite reliable sources, even when they have just been influenced by such arguments. Both issues should put a significant brake in the spread of political numeracy through the two-step flow process, in particular in non-dialogic contexts.

Keywords: political numeracy; two-step flow; argument evaluation; argument production; argument transmission
Political scientists have noted that citizens have, on average, low knowledge of politically relevant facts (Delli Carpini & Keeter, 1996). Moreover, on some issues citizens appear to be systematically innumerate, for instance in overestimating the share of the population that is foreign born (Sides & Citrin, 2007; Hopkins et al., 2019; Grigorieff et al., 2020), or the share of the budget spent on foreign aid (Kull, 1995). This ignorance is problematic at least at the collective level, since it might affect attitudes towards some of these political issues, which in turn could have an influence on policy-making (Gilens, 2001; Kuklinski et al. 2000).

Given the limits on the influence of mass media (see, Klapper, 1960; Mercier, 2020), this innumeracy might be attenuated by other sources of influence, in particular the ability of citizens to relay information to each other directly. In two-step flow models some citizens play a crucial role in relaying information from primary sources to the rest of the population (e.g., Huckfeldt et al., 1995; Katz & Lazarsfeld, 1955). In this article, we investigate two mechanisms that are crucial in the transmission of information between citizens: the willingness of informed citizens to use good arguments, and the ability of other citizens to recognize the quality of these arguments and change their minds accordingly.

We start by briefly reviewing the literature on two-step and multistep flow models, stressing the importance of argumentation in the process. We then turn to the two cognitive skills that are essential for the transmission of information through argumentation: the ability to update one’s beliefs when presented with good arguments,
and the ability to produce good arguments in order to convince others. Finally, we introduce our hypotheses and an overview of the present experiments.

**Two-step flow models: the importance of argumentation**

Political scientists and sociologists have noted that information doesn’t only flow directly from primary sources such as the mass media to the public. In two-step (or, later, multistep, see, e.g., Jensen, 2016) flow models, information transfer is mediated by informed citizens. For example, the 2014 World Values Survey suggests that over 40% of Americans get news daily from friends and colleagues (cited in Carlson, 2019). Moreover, political discussions between citizens do change minds (e.g., Goldberg et al., 2019).

To better understand what enables information transmission in two-step flow models, we can turn to the cognitive mechanisms with which people evaluate communicated information. As we explain presently, besides trust, the quality of the arguments presented matters a great deal in how people evaluate information (see, e.g., Petty & Wegener, 1999; Sperber et al., 2010).

Argumentation is critical in information transmission for at least four reasons. First, trust is not always sufficient to allow information transmission: even someone deemed more competent is not systematically believed (see, e.g., Moussaid et al., 2017). Second, an important factor in gauging expertise is the ability to provide good arguments. From a very young age (e.g., Castelain et al., 2018), we deem more competent those who provide good arguments. Third, arguments are easier to transmit than trust. For
example, if someone convinces you that the percentage of the budget spent on foreign aid is negligible by citing a reliable source, you should be able to use this source to convince someone else in turn. Fourth, not only are arguments potentially easier to carry along transmission chains, they are also easier to broadcast to large audiences. For instance, on social media, an individual’s expertise can only be fully recognized by their long-time followers. By contrast, any individual can make a good argument that reaches a wide audience, including people who had never heard of this individual.

Here, we focus on the role of argumentation in the transmission of politically relevant facts. More specifically, we focus on the role of the transmission of a single argument, by contrast with the back and forth of a discussion. The absence of back and forth with the audience is a feature of nearly all traditional media (Hampton et al., 2017), and the restriction to a short argument a feature of much of social media. As a result, the experimental conditions have at least a decent high ecological validity.

*Argument evaluation*

Much evidence shows that people are good at evaluating others’ arguments, even if these arguments challenge their pre-existing opinions. Participants consistently evaluate as being stronger, and are more influenced by arguments deemed superior by experts or by a normative framework such as Bayesianism (e.g., Hahn & Oaksford, 2007; Petty & Wegener, 1998; Chong & Druckman, 2007; on the efficacy of misinformation correction more generally, see, Walter et al., 2020; Bode & Vraga, 2018).
With respect to political issues, studies have also shown that presenting accurate information to people can change their minds on relevant policies (for instance, on how many people pay the estate tax, Sides, 2016). However, in these studies, the experimenters directly provide participants with information from a reliable source, assuming they will accept it. This is quite different from the situation in which a citizen attempts to convince another by citing a reliable source—the situation we explore here.

Even if people’s ability to be influenced by sound information and good arguments is clearly established, the literature also suggests that, when people are presented with an argument without the back and forth of an actual discussion, many remain unconvinced. This is true for arguments about numerical estimates (Minson et al. 2011), and even for logical arguments that would be nearly universally accepted during a discussion (Claïdière et al., 2017). Thus, the literature leads to the prediction that, when people receive a single argument supporting a numerical estimate, they should be more likely to accept the numerical estimates when the argument is good (rather than bad, or there being no argument), but that even good arguments should not convince everyone.

**Argument production**

As with almost any kind of information, arguments that are transmitted will often undergo significant changes in the process. These changes are due in part to the fact that people often produce, at first at least, arguments of mediocre quality. More specifically, the literature suggests that people are often not very exigent regarding the quality of the arguments they produce—which includes arguments inspired by others, and thus cases of argument transmission. People tend to spontaneously produce
mediocre, superficial arguments, for instance when asked to defend their views on social issues (Kuhn, 1991; Nisbett & Ross, 1980; Perkins, 1989). This trend is attenuated in dialogic contexts, in which people push each other to produce better arguments (Resnick et al., 1993). Still, the production of superficial arguments can make it difficult for informed citizens to transmit their knowledge: in the absence of feedback from their audience, these citizens might be content with providing relatively poor arguments to defend their numerate beliefs. We might expect that outside of dialogic contexts, even well-informed citizens often produce superficial, unconvincing arguments.

Within the skills of argument production, the most relevant for factual content is the ability to produce good arguments based on information one has already acquired, in particular on the basis of arguments received that cited reliable sources. We thus turn to experiments that have investigated argument transmission: people’s ability to produce good arguments after having been exposed to good arguments. Using transmission chain experiments, in which participants must pass arguments to each other consecutively, studies have found a wide range of outcomes, from a dramatic loss in argument quality (Bartlett, 1932), to near-perfect preservation of arguments (Claidièrè et al., 2017).

Studies on the transmission of politically relevant arguments in particular have revealed two main phenomena: loss, and bias. Carlson (2018) provided participants with information about an election, and asked them to pass it along, in a transmission chain; most information was gradually lost in the process of transmission. Other experiments observed that not all elements are equally likely to be lost. Aarøe and Petersen (2018) found that participants were more likely to transmit narratives rather
than statistical information. Boggild et al. (2020) observed that participants were more likely to transmit negative rather than positive information about politicians. Finally, Carlson (2019) found that participants were more likely to transmit information that fit with their political orientation, and that participants were less likely to be influenced by second-hand information than by information from a primary source (e.g., a news article). However, Carlson also found that a small subsample of participants could be as efficient at transmitting facts than the original news article.

**Hypotheses and overview of the experiments**

Based on the literature reviewed above, our two main hypotheses are:

H₁ For politically relevant facts, participants change their mind more when they are exposed to factual numerical information that is accompanied by good arguments (in particular, arguments citing a reliable source) than when it is accompanied by bad arguments or no arguments at all. This is true even when the arguments come from another participant and not an authoritative source, as in most previous experiments.

H₂ The majority of participants produce weak arguments to defend their opinions, but a small minority is able to produce good arguments by citing the reliable sources they used to form their opinion.

All our experiments use politically relevant numerical estimates, such as the percentage of the budget spent on foreign aid, or what share of the population pays the estate tax.
The experiments are organized as follows. In Experiments 1, 2, and 3 (testing H1.), participants are asked to answer questions about politically relevant facts, are provided with good arguments (i.e. citing reliable sources) supporting the correct answer, and given the opportunity to change their minds. The degree to which they change their minds is compared to different control conditions. Experiment 4 extends these results by testing the convincingness of a wider range of participant-generated arguments.

We then turn to H2. Experiment 5 is similar to Experiments 1 to 4, except that, after having provided their final answer—which can be influenced by the good argument they have just received—participants are asked to convince someone else in turn. Experiment 6 is similar to experiment 5, but makes it much easier for participants to provide an argument citing a reliable source. Experiments 5 and 6 are thus direct tests of the effectiveness of the two-step flow process.

**Experiments 1, 2, and 3**

In each experiment, participants are presented with a numerical question bearing on a politically relevant fact. These questions were pretested to ensure that most participants would provide estimates that significantly deviate from reality (see Electronic Supplementary Materials). After participants had given their initial estimates, they were provided with another answer (the advice), presented as coming from another participant (we rely here on the terminology used in advice-taking experiments, Bonaccio & Dalal, 2006; Yaniv, 2004). This answer was in some conditions supported by an argument. Following the presentation of the advice and (in some conditions) the argument, participants were given the opportunity to change their
minds. The flow of the experiment is described in Figure 1. In order to test the participants' ability to evaluate arguments, the advice and the arguments presented to the participants were manipulated across conditions.

### Initial Estimate

What percentage of the population in the United States today identify themselves as Muslims? If you do not know, you can just give your best guess.

### Advice Phase

To the question "What percentage of the population in the United States today identify themselves as Muslims?" you answered 14% [the participant's initial estimate]. The same question was answered by a previous participant, who answered: (...)

<table>
<thead>
<tr>
<th>Advice</th>
<th>Good Advice / Good Argument</th>
<th>Bad Advice / Bad Argument</th>
<th>Good Advice / Generic Argument</th>
<th>Good Advice / No Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(...) and gave the following justification:

- **According to the Pew Research Center, the Muslim population in the US is just a smidge over 1%, if you include children**
- **There aren't a lot of Muslims in the midwest but they usually live around the east coast**
- **I guessed this number based on my personal experience**
- **∅**

("and gave the following justification:" is removed)
### Final Estimate

What percentage of the population in the United States today identify themselves as Muslims? If you do not know, you can just give your best guess.

---

### Questions used

**Immigration:** What percentage of the United States population today would you say that was born outside the country?

**Religion:** What percentage of the population in the United States today identify themselves as Muslims?

**Foreign Aid:** Each year the United States government provides financial aid to foreign countries. What percentage of the federal budget do you think goes to foreign aid?

**Estate Tax:** In the United States there is a tax on large inheritances that is sometimes called the “estate tax.” What percentage of Americans do you think have a large enough estate to be subject to this tax?

**Pollution:** There is a lot of talk these days about environmental pollution from carbon dioxide (CO2). What is your best guess of the percentage of total CO2 emissions that comes from the United States?

**Education:** What percentage of the United States population today would you say have earned at least a four-year college degree - that is, a bachelor degree?
Figure 1. Experimental design of Experiments 1, 2, and 3 (Initial Estimate phase was skipped in Experiment 3), using as an example the question Religion, in the Good Advice / Good Argument, Bad Advice / Bad Argument, Good Advice / Generic Argument, Good Advice / No Argument condition, and list of all the questions used

Participants

We recruited 1199 US-located participants through Amazon Mechanical Turk (516 women; M Age = 34.08; SD = 10.8), paid $0.2. 600 participants took part in Experiment 1 (291 women; M Age = 35.37; SD = 11.02), 301 in Experiment 2 (113 women; M Age = 32.72; SD = 9.98), and 298 in Experiment 3 (112 women; M Age = 32.87; SD = 10.86). Approval from an ethics committee was obtained for all experiments (CER-Paris Descartes; No. 2019-03-MERCIER).

Design and procedure

After the consent form, Experiments 1 and 2 consisted of three phases, two of which (Initial Estimate and Final Estimate) were identical in both experiments (see Figure 1). In the Advice phase, participants were given a piece of advice and supporting argument, which were presented as coming from another participant. In order to test participants’ ability to evaluate arguments, the advice and argument they received were manipulated across six conditions of a between-participants design. Participants answered one question each.
In order to generate the advice and the arguments, we relied on a pretest which had the same structure as Experiments 1 and 2, except that participants were asked to give an argument to justify their final estimate, and they were incentivized to give good answers and good justifications (see Electronic Supplementary Material, henceforth ESM). We used some of these answers as advice and arguments in the present experiments.

In the Good Advice / Good Argument Condition (Experiment 1), participants received a good piece of advice (i.e. the correct estimate) and a good argument (i.e. an argument citing a specific and reliable source to support the estimate, see Figure 1 for an example, and ESM). Given that the present study focused on politically-loaded issues involving numerical information for which there is a correct answer that can be ascertained by consulting the relevant sources, we considered good arguments to be those that supported the answer with a reliable external source (e.g., Pew Research Center)\(^1\).

However, even if participants change their minds in this condition, they might do so not thanks to the quality of the argument, but, for instance, merely because they had very weak priors and would be willing to accept any piece of advice. To test this possibility, we introduced the following three conditions.

In the Bad Advice / Bad Argument Condition (Experiment 1), participants received a bad piece of advice and a bad argument (i.e. the argument that accompanied the bad answer in the pilot data, see Figure 1 for an example, and ESM). If participants recognize that the argument is poor, they should change their minds less than in the Good Advice / Good Argument condition. Since the advice also differs between these two conditions, 

\(^1\) All Experiments were conducted between Fall 2016 and Summer 2017.
it is possible that participants are less influenced in the Bad Advice / Bad Argument Condition because the advice in this condition is closer to their initial estimate.

The Good Advice / Generic Argument Condition (Experiment 1) controls for this by presenting the same advice as in the Good Advice/Good Argument Condition, but accompanied by a poor, non-committal argument ("I guessed this number based on my personal experience"). Note that a single argument was used in this condition. If participants place greater weight on argument quality, they should be less likely to adopt the advice in the Good Advice / Generic Argument Condition than in the Good Advice / Good Argument Condition. Still, one might suggest that this effect is mostly driven by participants being put off by the poor quality of the arguments in the Bad Argument and Generic Argument conditions, rather than the quality of the argument in the Good Argument condition.

In the last condition of Experiment 1, the Good Advice / No Argument Condition, participants are only presented with the good advice. If they change their minds less in this condition than in the Good Advice / Good Argument Condition, this will suggest that participants have recognized the strength of the good argument and changed their minds accordingly.

To summarize, $H_1$ will be supported if participants change their minds more in the Good Advice / Good Argument Condition than in any other condition.

To confirm the robustness of the change of mind in the Good Advice / Good Argument Condition, this condition was replicated in Experiment 2, and it was duplicated by a
Good Advice / Good Argument (2) Condition, which was identical to the former, except that a different good argument was selected for each question (see ESM).

It is possible that the provision of an initial estimate in Experiments 1 and 2 leads some participants to stick to this answer and discount the advice. In Experiment 3 participants do not provide an initial estimate before they are exposed to the advice. Two conditions are used, which are otherwise identical to the Good Advice / Good Argument Condition and Good Advice / Bad Argument Condition from Experiment 1. By comparing the Good Advice / Good Argument Condition in Experiment 1 (and its replication in Experiment 2) and Experiment 3, we can test whether the provision of an initial estimate dampens the degree to which participants are influenced by the advice.

Finally, in all three experiments participants answered demographic questions.

**Results**

To measure the extent to which participants shifted towards or away from the advice, we computed for each answer in Experiments 1 and 2 a Weight of Advice. Following Yaniv (2004) we take the Weight of Advice to be \(|f - i| / |a - i|\) (with \(f\) = final estimate; \(i\) = initial estimate; \(a\) = advice). This raises issues for a few categories of answers. If the initial estimate and the advice are equal, the Weight of Advice cannot be computed, and these cases were excluded from the analyses. If the final estimate goes beyond the advice, the Weight of Advice is equal to 1 (see, Gino, 2008, Gino & Moore, 2004, Soll & Larrick, 2009, and others). Finally, when the advice backfires, we add a negative sign in front of the computed Weight of Advice, and introduce a ceiling of -1 (to avoid potential
outliers and to maintain symmetry with cases in which the advice is overshot). As a result, the Weight of Advice ranges from -1 (maximum backfire effect) to +1 (the advice is adopted or overshot). To illustrate, consider a participant who provides an initial estimate of 14% to the question of the percentage of the US population who considers itself Muslim. Then, she receives a Good Advice (1%), which she either fully adopts in her final estimate (1%), adopts only partially (7%) or rejects (14%). The WoA for each of these scenarios, given her initial estimate and the Good Advice, would be equal to 1.00, 0.5 and 0.00, respectively.

The Weight of Advice was computed for 869 answers in Experiments 1 and 2, and it couldn’t be computed in 32 cases since their initial answers were identical to the advice. There were 17 cases of overshooting, and 26 cases of backfiring, including only one for which we replaced the computed Weight of Advice with -1.

The distribution of the Weight of Advice was strongly bimodal, with a majority of participants not changing their minds at all, and a plurality adopting the advice (Figure 2 and Table 1). This bimodal distribution was present across participants, regardless of their self-reported political affiliation (see ESM).

---

2 The distribution is thus far from being normal. However, since t-tests are very robust against non-normality, we have relied on them rather than on Wilcoxon tests (Rasch & Guiard, 2004). The same analyses using Wilcoxon signed-rank test can be found in the ESM.
Figure 2. Distribution of the densities of Weights of Advice in the three Good Advice/Good Argument conditions of Experiments 1 (first panel) and 2 (second and third panel)
<table>
<thead>
<tr>
<th>Condition</th>
<th>Weight of Advice</th>
<th>Distance Final Estimate – Advice</th>
<th>Percentage of participants who did not change their minds (WoA = 0)</th>
<th>Percentage of participants who changed their minds (WoA &gt; 0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good Advice/Good Argument</td>
<td>0.23</td>
<td>13.49</td>
<td>61.1%</td>
<td>35.1%</td>
</tr>
<tr>
<td>Bad Advice/Bad Argument</td>
<td>0.09</td>
<td>N/A</td>
<td>74.1%</td>
<td>18.9%</td>
</tr>
<tr>
<td>Good Advice/Generic Argument</td>
<td>0.15</td>
<td>16.2</td>
<td>70.9%</td>
<td>26.4%</td>
</tr>
<tr>
<td>Good Advice/No Argument</td>
<td>0.22</td>
<td>13.8</td>
<td>57.3%</td>
<td>39.9%</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good Advice/Good Argument</td>
<td>0.47</td>
<td>9.31</td>
<td>36.6%</td>
<td>62.1%</td>
</tr>
<tr>
<td>Good Advice/Good Argument (2)</td>
<td>0.52</td>
<td>9.62</td>
<td>31.7%</td>
<td>66.2</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good Advice/Good Argument</td>
<td>N/A</td>
<td>9.21</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Good Advice/Bad Argument</td>
<td>N/A</td>
<td>13.3</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 1.** Average Weight of Advice, average distance between final estimate and advice, percentage of participants who did not (WoA = 0) and did change their minds (WoA > 0.5), for each condition of Experiments 1, 2, and 3. No distance was computed for the Bad Advice/Bad Argument condition.
Argument condition of Experiment 1, since in this condition the advice was different from all the other conditions, the Weight of Advice could not be fairly compared. Finally, since no initial estimate was required in Experiment 3 no Weight of Advice could be computed.

To test H1, we compared the Weight of Advice in the Good Advice / Good Argument Condition, and the other conditions in Experiment 1. The Weight of Advice was significantly larger in the Good Advice / Good Argument Condition (M = 0.44, SD = 0.45) than in the Bad Advice / Bad Argument Condition (M = 0.087, SD = 0.299; t(252.31) = 7.94, p < 0.001, d = .92), b), the Good Advice / Generic Argument Condition (M = 0.15, SD = 0.29; t(247.56) = 6.6, p < 0.001, d = .76) and c), and the Good Advice / No Argument Condition (M = 0.22, SD= 0.33; t(267.39) = 4.695, p < 0.001, d = .55). All statistical tests reported were corrected for multiple comparisons using Bonferroni-corrected t-tests. Experiment 2 replicates the large change of mind observed in the Good Advice / Good Argument Condition of Experiment 1, and extends it to another set of arguments.

Given that no initial estimate was required in Experiment 3, the Weight of Advice could not be computed. We used the absolute distance between the final estimate and the advice, another measure of how much the participants took the advice into account. This distance was smaller when the arguments were good (Good Advice / Good Argument Condition; M = 5.03) than when they were bad (Good Advice / Bad Argument Condition; M = 13.3) (t(216.58) = -5.92, p < 0.001, d = .68).

Computing and aggregating the same distances for Experiments 1 and 2 allowed comparing the Good Advice / Good Argument conditions across the two designs (i.e.
with and without an initial estimate). Participants’ final estimate were closer to the advice when the participants had not provided an initial estimate (M = 5.03, SD = 7.39), than when they had (M = 9.94, SD = 15.2; t(510.46) = -5.176, p < 0.001, d = .41).

Overall participants benefited from being presented with good advice and good arguments, as shown by the significant reduction in average distance to the correct answer between the initial estimate (M = 17.84, SD = 17.33) and the final estimate (M = 9.94, SD = 15.20; t(853.19) = 7.15, p < 0.001, d = .48) in these conditions.

Discussion

Experiments 1, 2, and 3 show that participants change their minds more when presented with a numerical advice that is paired with a good argument (i.e. an argument citing reliable sources) than with no argument or with a poor argument, even when these arguments come from other participants instead of the experimenters. When they were confronted with a good argument, many participants moved towards the advice supported by this good argument, and approximately one quarter adopted the advice completely. Still, the modal answer, adopted by nearly at least a third of the participants, was to not take the advice into account at all, even when it was supported by a good argument.

Experiment 4

The main result from Experiments 1, 2, and 3 is that good arguments presented as coming from another participant prompt participants to substantially change their
minds on politically relevant facts. Experiment 4 extends these results by using a wider and more diverse pool of participant-generated arguments. All the arguments used support an approximately correct answer, but they are not all necessarily good.

**Participants**

We recruited 79 US-located participants through Amazon Mechanical Turk to generate the arguments (36 women; M Age = 35.6; SD = 11.75), and 411 for the main part of the experiment (181 women; M Age = 36.36; SD = 11.43).

**Argument generation**

To generate arguments that would then be provided to the participants, we recruited a first sample of 79 participants, who took part in an experiment broadly similar to the Good Advice / Good Argument Condition of Experiment 1, but in which they were asked to generate arguments for their final estimate. We collected the arguments of 41 participants whose estimates were close to the correct answer (at most four points difference) on the Immigration and Religion question (see ESM for details of the procedure).

The arguments generated were coded according to the following scheme. The objective was to isolate arguments that supported the estimate by providing a specific, reliable source (as do the good arguments in Experiments 1, 2, and 3).
- **Specific source (N = 19):** the argument mentions a specific, reliable source supporting the advice: “According to USA Today, the percentage of people living in the USA who were born outside the country reached 13.7% in 2015.”

- **Generic source (N = 13):** the argument mentions a source, but without providing a name for the source: “I have seen many news articles about Muslims recently and have seen that the population is under 5%”

- **No source (N = 9):** any other argument: “I don’t have a good justification. My answer was based and perception and mainly a guess.”

All answers were coded by one of the authors, and then recoded by a colleague (Kappa: 0.91). Disagreements were resolved through discussion.

**Design and procedure**

Experiment 4 had an identical structure to that of Experiments 1 and 2: participants provided an initial estimate to a factual question, they were provided with the answer and supporting argument of another participant, and could provide a final estimate (see Figure 1). The design of Experiment 4 differed in two respects. First, participants were asked to provide their estimate for only one of two questions: Immigration or Religion. Second, each participant was provided with one of the pieces of advice and argument generated specifically for this purpose (see above). Approximately 10 participants saw each participant-generated advice and argument.
Results and discussion

As above, we excluded 13 participants whose initial estimate was equal to the advice leaving 398 participants for whom the Weight of Advice could be computed (M = 0.33, SD = 0.42). The Weight of Advice was higher when participants had received an argument containing a specific source (M = 0.43, SD = 0.44), than a generic source (M = 0.31, SD = 0.38; t(295.24) = 2.59, p = 0.009, d = 0.29), or no source (M = 0.12; SD = 0.31; t(225.17) = 6.649, p < 0.001, d = 0.81). The Weight of Advice was also higher when participants received an argument containing a generic source than no source (t(203.6) = 3.98, p < 0.001, d = 0.54).

Discussion of Experiments 1 to 4

In Experiments 1 to 4, participants had to answer a politically relevant factual question (such as the percentage of the population born outside the country), were provided with arguments of varying quality, and offered the opportunity to change their minds (participants did not have to provide an initial estimate in Experiment 3). When presented with good arguments—arguments that mentioned a specific, reliable source in support of the numerical advice—approximately half of the participants moved towards the advice, and a quarter fully adopted it. By contrast, when they were presented with no argument, or with weaker arguments, fewer participants changed their minds.

Crucially, all the arguments used in Experiments 1 to 4 were presented as coming from another participant, instead of a source that is itself authoritative (such as the
experimenter). In terms of two-step or multistep flow models, this means that if citizens are able and willing to muster good arguments, then they should be able to exert some influence on other citizens.

In Experiments 1 to 4, the arguments we gave the participants had been generated by other participants, who had been incentivized to produce such arguments. These experiments are thus not suitable to measure the ability and willingness of the participants to provide such arguments in the absence of such incentives. In Experiments 5 and 6, incentives were removed, allowing us to test whether participants would spontaneously produce good arguments to defend their opinions.

**Experiment 5**

As argued in the introduction, for two-step or multistep flow models to function, it is likely crucial that better informed citizens use good arguments to defend their opinions. The goal of Experiment 5 is to test whether participants who have been exposed to a reliable source of information are willing to use that source of information when convincing someone else in turn. Experiment 5 is similar to Experiments 1, 2, and 4 in that participants are asked to give an initial estimate to a factual question, followed by the provision of a correct answer supported by a good argument. Unlike our previous experiments, where the argument was presented as coming from another participant, in Experiment 5 the source is merely stated, and thus—as in nearly all past studies—implicitly endorsed by the experimenters. Then, after they have provided a final estimate in light of this information, participants are asked to generate a reason that could convince another participant to accept this answer. Experiment 5 thus puts
participants in the potential position of informed citizens who have received correct information from a reliable source, and who are asked to pass on that information to others.

Participants

We recruited 202 US-located participants through Prolific Academic (122 women; M Age = 32.22; SD = 36), paid £1.

Design and procedure

After agreeing to participate in the study, participants were presented with six different politically relevant factual questions, in a randomized order. Then, participants were provided with the correct answer, supported by a specific, reliable source. At this stage, participants were given the opportunity to change their initial estimate and asked to justify “so that it might convince someone who hasn’t seen the answer that was provided to [them].”

Results

We excluded 23 answers for which participants’ initial estimates were equal to the advice provided. The average Weight of Advice for the remaining answers was 0.38 (SD = 0.46).
Of particular interest were those answers in which participants changed their mind in response to the advice, and thus can be expected to have paid some attention to the argument. We defined the answers in which participants changed their minds in response to the advice as those answers in which (i) the initial estimate was sufficiently distant from the advice to allow for a meaningful movement towards the advice (i.e. a difference of at least five points) and, (ii) the final estimate had moved all the way or nearly all the way towards the advice (i.e. Weight of Advice of at least 0.9).

A total of 267 answers reflected such a change of mind (see Table 2, Experiment 5, row “Adopts the advice”, column “Total Answers”). All answers were coded by one of the authors, and then 20% were recoded by the other author (Kappa: 0.76).
Table 2. Number (and percentage) of individual answers providing each of the three categories of arguments in Experiment 5 (participants provided six arguments each) and Experiment 6 (participants provided one argument), for all participants, participants who did not adopt the advice, and participants who did (see text for details).

<table>
<thead>
<tr>
<th>Adopts the advice</th>
<th>24 (24%)</th>
<th>37 (37%)</th>
<th>39 (39%)</th>
<th>100</th>
</tr>
</thead>
</table>

To test whether participants who provided no source, or only a generic source, were simply unwilling to put any effort in the argument production task, we compared the number of words in each of these categories of arguments. Arguments relying on a specific source tended to have more words ($M = 28.1; SD = 18.8$) than arguments citing no source ($M = 20; SD = 16.4$; $t(78.51) = 3.56, p < 0.001, d = .43$), but not that arguments citing a generic source ($M = 28.8; SD = 23.2$; $t(121.55) = -0.19, p = 0.85, d = .03$).

Finally, to test whether the fact of merely stating the source, rather than presenting it as coming from another participant, had an effect on the convincingness of the arguments, we compared the Weight of Advice of Experiments 1 and 5 (Good Advice / Good Argument condition). This difference was not significant; $t(183.5) = 1.37, p = .17, d = .13$.

**Discussion**

In all the experiments so far, of the participants exposed to a good argument—citing a reliable source—about a politically relevant fact, approximately half changed their
minds. However, when, in Experiment 5, participants were asked to convince someone else in turn, they cited the specific source they had been exposed to in only 6% of their arguments. Even among the answers that had been heavily influenced by the argument, only 13% were supported by arguments citing a specific reliable source—the very source that had led to the answer. This is true even though participants who did not mention any source in their arguments still took the time to write down arguments averaging 20 words in length: shorter than the average 28 words of those mentioning a specific source, but still representing some effort, for a likely much lower degree of convincingness (as suggested by Experiment 4).

**Experiment 6**

Experiment 5 suggests that very few participants, even among those who have been influenced by an argument citing a reliable source, use this reliable source when they have to convince others in turn. We can consider this to be a lower bound to participants’ preference for citing reliable sources. In order to provide a potential upper bound (at least in the present, online setting), and to better understand why most participants failed to use the reliable source in their arguments in Experiment 5, we conducted a new experiment that was broadly similar to Experiment 5, but with some changes aimed at making it easier for participants to cite a reliable source. More specifically, in Experiment 6, the arguments that the participants had received were still displayed when they had to write down their own arguments, thereby removing any performance factor.

**Participants**
We recruited 603 US-located participants through Prolific Academic (301 women; M Age = 31.51; SD = 10.95), paid £0.38.

Design and procedure

As in Experiments 1, 2, and 4, participants were asked to answer a question about a politically relevant fact, provided with the correct answer of another participant, and an argument supporting this answer that relied on a specific reliable source. Participants could change their mind on the basis of this advice. As in Experiment 5, they were then asked to provide an argument supporting their final estimate.

Unlike Experiment 5, the arguments that the participants had received were still displayed when they had to write down their own arguments. Second, this argument was presented as coming from another participant, instead of from the experimenter. Our results however suggest that this is unlikely to have a significant effect on the argument’s convincingness, but it might reassure participants that providing such arguments is acceptable. Third, we only asked participants to answer one question (and thus recruited a larger sample), to avoid any decrease in motivation with repeated requests to justify their answers. Fourth, we only used two topics for the questions, Immigration and Religion (as in Experiment 4, and for the same reason: maximizing the share of participants who would change their mind on the basis of the advice). Fifth, and finally, to ensure that participants who had been influenced by the advice, and had thus given an at least approximately correct answer, would realize they needed an argument to convince others, we gave them the following prompt:
We are now asking you to write, in your turn, a justification for your answer. This argument will be passed on to another participant, along with your estimate. As an indication, the average answer to this question for people who have not seen another participant’s argument is about [average of answers from previous studies] %.

Results

We excluded nine participants whose initial estimates were equal to the advice provided. The average Weight of Advice for the remaining 594 participants was 0.47 (SD = 0.42) (Table 2, Experiment 6). All answers were coded according to the same scheme as in Experiments 4 and 5 by one of the authors, and then a subsample of 20% was recoded by the other author (Kappa: 0.82).

As in Experiment 5, we then split the participants into those who had changed their mind in response to the advice, and those who hadn’t (Table 2).

Finally, we found that those who changed their minds, compared to those who didn’t, were more likely to cite a specific source, compared to the other two argument categories (two categories collapsed: $X^2 (1, N = 594) = 38.46, p < 0.001$; two categories separated: $X^2 (2, N = 594) = 61.2, p < 0.001$).

Overall, participants wrote more words in their arguments when they mentioned the specific source ($M = 26, SD = 15.1$), compared to when they mentioned a generic source.
(M = 25.1, SD = 15.3), or no source at all (M = 23.4, SD = 17.3). None of these differences were statistically significant.

**Discussion of Experiments 5 and 6**

In Experiments 5 and 6, participants were provided with an argument citing a reliable source about a politically relevant fact. After they had had the opportunity to take this argument into account, they were asked to produce an argument in turn. The main difference between the two experiments is that in Experiment 6, participants could still see the argument citing the reliable source when they had to write their own argument.

Some participants were not influenced by the good argument—either because they had provided the correct answer before being exposed to the good argument, or because they stuck with their initially incorrect estimate. Among those participants, unsurprisingly, few used the reliable source they had been exposed to in their own arguments (Experiment 5: 3.9%, Experiment 6: 13%).

More surprisingly, even among those who had substantially changed their minds after accepting the answer supported by the good argument, when they had to to convince others in turn, only a minority made use of the reliable source that had convinced them. In Experiment 5, 13.5% of the arguments cited a reliable source, and even in Experiment 6, in which the good argument with its reliable source was available as participants typed in their own arguments, only 39% of the arguments participants generated cited the reliable source. These results suggest that our participants had, on
the whole, only a modest inclination to cite reliable sources in their arguments, even when doing so would have arguably been the easiest option for them.

**Conclusion**

Political scientists stress the importance of interpersonal communication in the spread of ideas, by contrast with direct messaging from the media or other institutional sources. When the ideas being communicated don’t agree with the audience’s prior beliefs, arguments should play a crucial role in helping these ideas spread. In this article, we tested whether arguments could play such a role in the spread of knowledge of politically relevant facts.

The literature suggests that people are able, on the whole, to evaluate arguments relatively well, even when the arguments challenge their priors. However, when arguments are presented outside of a dialogic setting, they seem to lose much of their effectiveness, and some people remain unconvinced. The literature also suggests that, when it comes to argument production, when asked to defend their views outside of dialogic contexts, people do not put in much effort, producing somewhat shallow arguments. These trends are mirrored in the present results.

In our experiments, we asked participants about politically relevant estimates for which there are widespread misconceptions (Grigorieff et al., 2020). For instance, the share of US citizens paying the estate tax is 0.2%, while our participants (Experiment 1) answered 21% on average. After having asked participants to give an initial estimate, we provided them with a good advice supported by a good argument, citing a reliable
source such as the Pew research center or the Census bureau. Participants could then revise their views, producing a final estimate. In all but Experiment 5, the arguments participants saw had been produced by other participants, and were presented as such (in Experiment 5, the arguments came directly from the experimenters). Finally, in Experiments 1, 4, 5 and 6, participants were then asked to defend their final estimates.

Across all the experiments, we observed that participants presented with good arguments moved on average nearly half of the way towards the correct answer (unweighted mean Weigh of Advice = 0.45)—although it is more accurate to say that many participants adopted the correct answer, while many others rejected it altogether. These experiments on argument evaluation add to the previous literature in two ways.

First, the arguments citing the reliable source came from another participant instead of an experimenter. This shows that people focus on the quality of the argument, not its source, and that if a citizen cites a reliable source, they have reasonable odds of convincing a peer.

Second, we included a number of control conditions in which participants were either provided with the correct answer not supported by any arguments, the correct answer supported by poor arguments, or an incorrect answer supported by poor arguments. In each of these conditions, participants were much less influenced by the answer they had received, suggesting that good arguments are crucial for people to change their minds.

In Experiments 5 and 6, participants were asked to justify the answers they had given after having been provided with the correct answer and a good argument citing a
reliable source. Of particular interest are participants who were heavily influenced by the argument: those whose initial estimates were wrong, and yet who accepted the correct answer after reading the argument. For these participants, we can be confident that they owe their final estimates to the exposition to the correct answer and its supporting argument. These participants should also realize the persuasiveness of the argument they have seen, since it has just convinced them. However, only a very small minority (13% of answers in Experiment 5) of these participants mentioned a reliable source in their own arguments.

In Experiment 6, the good argument and its reliable source were still displayed when the participants had to type their own argument. Yet most of the participants who had been convinced by the good argument failed to recycle its reliable source in their own arguments (only 39% did so). This result suggests that the failure of most participants to cite a reliable source in their arguments isn’t a failure of memory, but one of motivation: they appear not to find it relevant to cite a reliable source in order to convince someone else to accept an answer. Moreover, the results of Experiment 4 suggest that, by failing to mention reliable sources in their arguments, participants lost much in terms of convincingness.

Overall, our results are reminiscent of those of Carlson (2019) who had found that only a small share of participants could efficiently transmit politically relevant news.

Why would so many participants choose not to cite a reliable source in order to convince others, even when such a source is easily accessible? Curiously, the same pattern has been observed in at least two other very different settings (Claidière et al.,
It is thus conceivable that there is a common reluctance to mention that we owe ideas we believe to be correct to others, even if that means failing to convince them. A tentative explanation for this pattern is that people gain more credit for good ideas if the ideas are presented as being their own, rather than something they owe to someone else (Altay et al., 2020). Even in the case of political estimates, which one can hardly discover on one’s own, there might be more credit to be gained by claiming these facts as something one ‘just knows,’ instead of something one has just found out by reading an article. Similarly, recent research has shown that when it comes to moral and political issues, subjective experiences are doubted less than objective facts (Kubin et al., 2021). This might partly explain participants’ reluctance to mention reliable, “objective” sources in their own arguments.

Irrespective of the reason why people appear reluctant to mention a reliable source in their arguments, this reluctance should hinder the spread of accurate beliefs. In our experiments, few participants changed their minds if they were not exposed to a good argument. If even the citizens who have acquired accurate beliefs through exposure to reliable sources do not use these sources in their arguments, their accurate beliefs will not spread much further.

Our study has several limitations. First, our samples were not representative (although that does not seem to have a huge impact on experimental manipulations, see, Coppock, 2018). Second, we did not measure whether participants were really misinformed about the issues (as our results appear to suggest), or merely uninformed (see, Li & Wagner, 2020). This distinction might help understand why some participants changed their minds dramatically in response to the advice, and others didn’t at all. The last and
arguably most important limitation is that the experimental context might not have been conducive to the production of good arguments. By contrast with our setting, in a discussion participants’ poor arguments should be rejected, which might prompt them to produce better arguments, in particular arguments citing reliable sources. However, in many contexts people do not get the feedback that would prompt them to produce better arguments, and, as in our experiments, they might get stuck on poor arguments that do not convince anyone (as the comment section on most newspaper articles attests).

With these preliminary studies, we hope to draw attention to potential difficulties in the spread of accurate beliefs, difficulties due in part to some citizens’ reluctance to change their minds even when confronted with good arguments, and, potentially, to the lack of motivation to cite reliable sources in order to convince others of the accurate beliefs they hold. Given that the latter phenomenon has been less studied, future research should investigate its limits, testing for instance whether it is attenuated in dialogic contexts, or when people have had time to recognize expertise and built trust in each other.
Acknowledgments

This work was supported by the French National Research Agency (ANR) under grants ANR-17-EURE-0017 FrontCog and ANR-10-IDEX-0001-02 PSL.

Disclosure statement

The authors report there are no competing interests to declare.
References


