# Improving deliberations by reducing misrepresentation effects

Cyrille Imbert*[1], Thomas Boyer-Kassem†[2], Vincent Chevrier[3], and Christine Bourjot[3]

[1]Archives Poincaré, CNRS, Université de Lorraine, Nancy, France
[2]MAPP (EA 2626), Université de Poitiers, France
[3]Université de Lorraine, CNRS, LORIA, France

Forthcoming in *Episteme*.

Submitted: 8 February 2018; revised: 15 April 2018; accepted: 20 July 2018

***Key words***— deliberation, deliberative procedures, misrepresentation, deliberative groups, epistemic groups, social epistemology, formal epistemology, political epistemology, political philosophy, random order, computer simulations

**Abstract**

Deliberative and decisional groups play crucial roles in most aspects of social life. But it is not obvious how to organize these groups and various socio-cognitive mechanisms can spoil debates and decisions. In this paper we focus on one such important mechanism: the misrepresentation of views, i.e. when agents express views that are aligned with those already expressed, and which differ from their private opinions. We introduce a model to analyze the extent to which this behavioral pattern can warp deliberations and distort the decisions that are finally taken. We identify types of situations in which misrepresentation can have major effects and investigate how to reduce these effects by adopting appropriate deliberative procedures. We discuss the beneficial effects of (i) holding a sufficient number of rounds of expression of views; (ii) choosing an appropriate order of speech, typically a random one; (iii) rendering the deliberation dissenter-friendly; (iv) having agents express fined-grained views. These applicable procedures help improve deliberations because they dampen conformist behavior, give epistemic minorities more opportunities to be heard, and reduce the number of cases in which an inadequate consensus or majority develops.

## Acknowledgments

We would like to thank participants of various workshops held in Clermont-Ferrand, Munich, Nancy, Paris, Rotterdam, and Tilburg, in particular John Beatty, Denis Bonnay, Franz Dietrich, Hélène Landemore, and Philippe Urfalino, for valuable comments and suggestions. This work has benefited from financial support from CPER Ariane, Université de Lorraine, and an AXA postdoctoral fellowship.

*Corresponding author: `Cyrille.Imbert@univ-lorraine.fr`
†Corresponding author: `thomas.boyer.kassem@univ-poitiers.fr`

# 1 Introduction

Deliberative and decisional groups play crucial roles in most aspects of social life, from hiring committees, political groups, lay juries, and appeal courts to managerial boards of companies, non-profit organizations, or governmental agencies. However, there is growing evidence that it is not easy in such groups to carry out beneficial exchanges in which all relevant information and views are properly taken into account. Indeed, deliberative groups exhibit patterns such as extremism, homogenization, polarization, information cascades, or inability to pool information (see e.g. Kuran & Sunstein, 1998, Sunstein, 2000, Luskin, Fishkin & Hahn, 2007, Sunstein, 2010). Such factors feed the suspicion that their decisions could be distorted by social mechanisms. Accordingly, understanding how such features arise and how group deliberations can be improved is a valuable goal.

Such investigations are actively carried out in several fields. Social epistemology analyzes the epistemic value of social practices and the nature of epistemic groups and communities (Goldman, 1999, Lackey, 2014, List and Pettit, 2011, Martini and Boumans, 2014). These issues are also of major significance for political philosophy and political epistemology. Finding epistemically virtuous procedures is in particular crucial for epistemic democrats who want to justify the value of democracy on epistemic grounds (Cohen, 1986, List and Goodin, 2001, Estlund, 2008, Landemore, 2013). Finally, the epistemic performances of collectives, from collaborative groups and expert committees to scientific communities is also central in science, and epistemic influence within such collectives is intensively investigated by formal philosophers of science (see e.g. Mayo-Wilson, 2014, Zollman, 2010a, 2010b, 2012). Overall, finding appropriate ways of deliberating so that good informed decisions are finally made is a legitimate concern.

Some aspects of the processes that lead to group decisions have already been investigated, others less so. Procedures for choosing members of a decisional group, eligibility conditions, publicity and secrecy concerning deliberations, or voting procedures are important issues that have attracted much attention. Because deliberative procedures may have a strong impact on decisions, they are also worth investigating. This is a difficult and highly multi-dimensional issue since the epistemic value of deliberative procedures is partly sensitive to the cognitive and social behavior of deliberators.

In this paper, we analyze a potentially powerful mechanism which can alter the quality of deliberations. It is a documented fact that the public views of agents often differ from their private ones (Asch, 1951, Kuran, 1995, Hogg, 2010, Cialdini & Griskevicius, 2010). We analyze in this paper the effects of this type of conformism in groups carrying out sequential oral deliberations or votes, the ultimate objective being to identify deliberative procedures that could decrease the effects of this conformity mechanism.

We first present misrepresentation and emphasize that it is a widespread behavior, even among experts (Section 2). We further present a simple mathematical model and justify why it is appropriate for our investigation (Section 3). This models helps us to highlight the significant effects of misrepresentation and discuss the efficiency of different procedures to dampen them: (i) holding a sufficient number of rounds of expression of views (Section 4); (ii) choosing appropriate orders of speech (Section 5); (iii) rendering the deliberation dissenter-friendly (Section 6); (iv) encouraging agents to express fine-grained views (Section 7). We discuss in Section 8 the robustness and scope of our findings.

# 2 Deliberation and misrepresentation

## 2.1 Misrepresenting one's views

Agents may misrepresent their own views in various circumstances. When choosing a restaurant with friends, if others have already concurred on a particular venue, you may spontaneously misrepresent your preference so as not to spoil the party. In other cases, dissension comes with high costs, and incentives to misrepresent one's views are extremely strong. During the Christian re-conquest, it was safer not to be a publicly practicing Jew or Muslim in Spain. Similarly, coming out is socially easier for homosexuals when many others have already done so.

The economist Timur Kuran describes these situations, where agents do not make their genuine preferences public "under perceived social pressures", as cases of "preference falsification" (Kuran, 1995, 3). In other words, agents anticipate that their public behavior may come with major costs, and feel that dissenting would be too harmful for them. Consequently, they decide to bend their public preferences or behavior towards those of others. Preference falsification is thus a special case of conformist behavior, where agents' private preferences differ from those that they publicly endorse because of perceived pressures. It differs from cases in which agents mask their disagreement because dissenting voices have already been expressed in preliminary deliberations, agents have as a group jointly accepted a position, and they have agreed not to speak *in propria persona* publicly (Gilbert, 1987, Beatty, 2006).

Beyond preferences, falsification can also affect opinions, id est statements for which there are matters of facts and which can be true or false. To encompass both cases, we shall talk of "views". Further, instead of "falsification", we shall use the term "misrepresentation" which has no truth-related connotation and applies better to all kinds of views.

Misrepresentation can stem from various factors or mechanisms, such as conflict aversion, a desire to avoid being publicly wrong, a reluctance to appear publicly as a lone dissenter or a non-conformist, an unwillingness to have to defend actively and probably uselessly one's views against the majority, the fear of having to suffer from future retaliation from other people, or the belief that making an inaccurate statement will be beneficial[1]. Here, we shall not investigate these specific individual reasons for misrepresenting one's views but merely focus on the effects of this behavior on group deliberations.

## 2.2 The scope of the mechanism of misrepresentation

Intuitively, misrepresentation can be expected to be a major phenomenon when agents are uncertain about their views, or about issues pertaining to their social identity and for which social costs can be high, such as sexual orientation, religious beliefs, or political opinions. Surprisingly, such conformity effects can be strong when nothing serious is at stake, people have no specific reasons to fear other people in the group, and uncertainty

[1] In scientific contexts, fraud in reporting experimental results can be seen as a type of misrepresentation. However, in contrast with cases like those described by Kuran, scientific fraud does not seem to stem from compliance with social pressure, since it is usually rooted in personal motivations such as the desire to agree with trendy research and earn credit, or, paradoxically, the desire to promote truth in the case of "noble lies" (see Bright, 2017 for more detailed analyses).

is low. In Asch's experiment (Asch, 1951), agents were asked to carry out a trivial task (visual comparison of lines) for which less than 1% of people usually gave a wrong answer. But when subjects heard other people (actually, actors) around them unanimously give a wrong answer, in up to 32% of cases, they gave the same wrong answer.

Evidence seems to show that misrepresentation can also exist in groups of experts, when reputational concerns towards (academic) peers are important, power relations are present, and uncertainty about debated issues run high. Schlesinger, who in 1961 was Special Assistant to the President Kennedy, reported that, though strongly opposed to the Cuba's invasion, when he was faced in the cabinet room with an "intimidating group" (in favor of invasion) he "shrank into a chair at the far end of the table and listened in silence" (Schlesinger, 2002, p.240). Sunstein investigated the case of appeal courts in the USA from 1980 to 2002 (Sunstein, 2005). In these courts, judges tend to share views with the political party of the president who appointed them. A qualitative finding is that a judge's ideological tendency can be significantly dampened when she is sitting with judges appointed by a president of the other political party (ibid, p.168). Various potentially coexisting mechanisms probably contribute to this phenomenon, including the power of contradictory arguments, the search for a consensus, or the need for "collegial concurrence" (p.182). Still, as noted by Sunstein, misrepresentation mechanisms, like in Asch's experiment, also seem to be at work. More generally, it is well-known that experts, like ordinary people, fall prey to cognitive biases such as overconfidence or confirmation bias, so assuming that misrepresentation and conformist behavior does not touch them would be a bold bet. Indeed, academics care about their scientific reputation and may be reluctant to appear as unorthodox members in their community. Science is for a large part a conservative and communal activity (Kuhn, 1962), in which dissenters are not often rewarded. Overall, if misrepresentation can occur in the absence of uncertainty and of group pressure, while being present in expert groups, it can be expected to affect most deliberating groups.

This influence can alter the successive opinions that are presented in deliberations but also distort final decisions, especially if they are taken by oral votes. After a first phase of oral deliberation, a group may misleadingly think that it has reached a state of genuine consensus or unanimity, and that it is unnecessary, if not inappropriate, to formally proceed to a secret vote, thinking "We all agree, no need to vote, right?" As we shall see, notwithstanding appearances, this public consensus may be the result of a misrepresentation cascade[2]. Importantly, in such cases, secret voting is not a panacea. First, for pragmatic reasons: secret voting is a formal procedure that takes time and requires material, which makes the deliberation process laborious. Not all groups may be willing to pay this cost for all decisions. Also, there are cases in which a secret vote is not suitable for theoretical reasons (see Elster, 2015, for detailed discussions). For example, public votes are seen as requiring public justifications based on arguments and acceptable evidence, whereas secret voting is seen as opening a space for personal interests or biases. Thus, the demand for openness and accountability may require that agents express their views publicly. This is *de facto* a requisite in many scientific expert panels. For instance, the rules at the European Food Safety Agency, the European Environment

[2]See also Beatty and Moore (2010) for analyses about inadequate consensus and Kosolosky and Van Bouwel (2014) for a critical review of consensus-making.

Agency, the European Chemicals Agency, the European Medicines Agency, the committees advising the European Commission, the International Accounting Standards Board, make no mention of secret voting when it comes to scientific decisions, and can require that names be associated with opinions. In brief, public voting is not always dispensable and the corresponding decisions can also be warped by the effects of misrepresentation, which is strong ground for fine-tuning the deliberation and voting procedures of the corresponding institutions[3].

### 2.3 Misrepresentation and its effects

Publicly complying with others' behavior is not necessarily harmful. For individuals, conformism protects them from the dangers that minorities suffer. Also, others' behavior can be an informational signal telling them which choice is the best, or which claim is true; so following the group can be epistemically beneficial (Banerjee 1992). For groups, conformity or compliance can also promote consensus, cohesion, quicker decision-making, and sometimes reliability (Zollman, 2010, 2012).

However, opinion misrepresentation also brings with it detrimental effects. When individuals misrepresent their views, then the final group's view will be farther from theirs. For the group, the views of some individuals remain hidden. This potentially reduces the diversity of expressed opinions, which has been argued to be epistemically valuable (Mill, 1859, Landemore, 2013), and the group may be deprived of valuable information. If the decision concerns preferences, the collective choice may be less representative of the preferences of its individual members. In brief, collective deliberations should be potentially sensitive and responsive to the views of each of their members, all of which are potentially valuable, otherwise, why deliberate in the first place? The misrepresentation of views goes against this requirement.

The effects of misrepresentation can also be indirect and delayed. If, in a sequential deliberation, all agents misrepresent their views based on what they have already heard, cascades and snowball dynamics can be expected, with more weight being given to the first agents. Thus, a group may finally endorse a view that significantly departs from that of the vast majority of its members.

Overall, because of the importance of its effects, it is necessary to investigate further the impact of the misrepresentations of views on epistemic deliberations. Our first goal is to investigate how large and distorting these cumulated effects can be for deliberative groups. Our second goal is an analytic one; to understand how such cascade effects work in the present case and which factors favor them. Our third goal is both applied and normative: when the drawbacks of misrepresentation are seen as more important than its advantages, which applicable deliberative procedures could be used to decrease these effects?

[3]See for example (Urfalino and Costa, 2015) for an analysis of the 2007 reform of FDA Advisory Committee decision-making and voting procedures.

# 3 A simple model of misrepresentation

To reach our goals, we introduced a formal model of the problem and investigated it by computer simulations. Accessing and keeping precise track of the private views of all agents during a real group deliberation *without perturbing it* is difficult. Simulating a model enables us to sidestep this problem: the modeler can set the private opinions which are held, follow the details of views throughout the deliberation, make a systematic investigation for various parameter values, and assess by computational means the epistemic value of practices (Douven, 2009, Olsson, 2011).

During deliberation, the views expressed by certain agents may lead others to misrepresent their private views, but also to change those private views, so the misrepresentation mechanism could ultimately be combined with some model of opinion dynamics (e.g. Lehrer and Wagner, 1981, Deffuant et al., 2000, Hegselmann and Krause 2002, Weisbuch et al. 2005, Zollman, 2012). Here, we chose not to effect this combination but to investigate a model without dynamics of private opinion, where only the agents' public views could change. From an analytical perspective, starting by studying the separate effects of mechanisms is informational. This paves the way to understand complex cases in which several effects are entangled.[4] Also, with this modeling choice, all changes in agents' public views are the sheer effect of the misrepresentation mechanism. We thereby aim to reach global insights about how the distorting impact of misrepresentation depends on various parameters, not to make precise predictions about this impact in specific situations. In particular, we do not make commitments about what exactly takes place additionally when the misrepresentation mechanism feeds the actual epistemic dynamics of private opinions.

## 3.1 Private views and public views

Our model considers groups of $n$ agents (hereafter called $n$-groups), which must choose between two options, typically to adopt or reject a policy or a candidate. For instance, they may need to answer the question "Should abortion be authorized until the third month of pregnancy?" or "Should this drug be given marketing authorization?" Although the choice is binary, the views of agents may be actually richer than "yes" or "no". For instance, an agent may think "Yes, this is the least", or "hardly more yes than no". This is modeled by assuming that agent $i$ has a private opinion $p_i$ within $[0, 1[$, drawn from a flat $[0, 1[$ distribution. The interval, $[0, 0.5[$ (resp. $[0.5, 1[$) represents a view that is compatible with the first (resp. second) option, which is chosen to be 0.25 (resp. 0.75). That is, if the private view $p_i$ of agent $i$ is equal to or below 0.5, he or she favors option 0.25, and 0.75 otherwise.

During deliberations, agents publicly express binary views, i.e. 0.25 or 0.75 and not a more refined position between 0 and 1 (this presumption is relaxed in Section 7). Indeed, presenting exactly the details of a precise position on a complex issue is not straightforward. Agents may fail to express the subtleties of their views, the audience may not understand these subtleties or merely be interested in the option that individuals support. In brief,

[4]Given the existing wealth of opinion dynamics models (see e.g. Muldoon, 2012 for a philosophical review), analyzing misrepresentation effects in the framework of one such particular model would also render our results too specific.

without misrepresentation, agent $i$ would publicly express the view

$$e_i = \text{Proj}(p_i) \tag{1}$$

in $\{0.25, 0.75\}$, which is the projection of her private view on the closest available option. The expressed views are the only ones that are accessible to other members of the group.

With the choice of 0.25 and 0.75 as possible options (versus the choice of the extreme views 0 and 1), the two intervals $[0, 0.5[$ and $[0.5, 1[$ are symmetrically divided; agents may hold private views that are more extreme than the ones they can express. Also, the difference between a private and a public view which is due to the loss of information in the expression process cannot be greater than 0.25. Finally, this choice corresponds to a moderate misrepresentation mechanism: an agent who believes 0.45, and should say 0.25 but misrepresents his or her view because of social influence and says 0.75, misrepresents less than if the other option was 1. Accordingly, if we observe major misrepresentation effects, they will not come from a too favorable modeling assumption.

### 3.2 Misrepresentation

Including the misrepresentation of views in this model bends an agent's expressed view towards those already expressed. Agents voice their views publicly in a sequential way. During the first roundtable the first agent expresses her private view. When $i-1$ agents have already spoken and agent $i$ is about to speak (with $i > 1$), we define the group's expressed view $G_i$ as the linear average of the $i-1$ voiced views. After the first round, for an $n$-group, the group's expressed opinion $G_i$ for agent $i$ is defined as the linear average of the last $n-1$ expressed views, i.e. the average bearing on all other agents.

Because the linear average that defines $G_i$ is not weighted, agents give equal attention to the previous views expressed by all other agents. However, because of cascading effects, the earlier agents speak, the more their private views influence their peers. The first speaker even speaks his or her mind freely, initializing the group's expressed view and thereby influencing all subsequent speakers. This reflects a feature of actual deliberation: the first positions can be strategic in that they may start an opinion wave that is difficult to oppose. Expressing $A$ is easier if you speak first than if you speak after 4 unanimous speakers who have defended $not - A$.

We introduce a parameter $\alpha \in [0, 1]$ which represents how strong the misrepresentation mechanism is. Then, an agent expresses a view which lies in-between $p_i$, i.e. his or her private view, and $G_i$, what the group's view is for him or her. Accordingly, instead of expressing $\text{Proj}(p_i)$, for $i > 1$, agent $i$ eventually expresses

$$e_i = \text{Proj}((1-\alpha)p_i + \alpha G_i) \tag{2}$$

(and $e_1 = \text{Proj}(p_1)$). That is, he or she considers a weighted linear average of his or her private view and of the group's expressed view, and projects it onto 0.25 or 0.75. As indicated above, misrepresentation can be produced by different compatible mechanisms, and the parameter $\alpha$ aggregates them. For simplicity, the misrepresentation rate $\alpha$ is kept constant, but this assumption is relaxed in Section 8.

To study the effects of the misrepresentation of views, we shall compare the results of an oral sequential vote based on what agents have publicly expressed in each roundtable

and of a secret vote, which is equivalent here to an oral sequential vote without misrepresentation. In both cases, we assume that the group's decision is taken by applying the majority rule, which is frequently used in committees and governmental agencies. Because agents misrepresent their views, the deliberation may be warped, and these two votes can give different results. We call these majority group decisions that are reversed because agents publicly misrepresent their views "distorted" decisions. We explore by computer simulations how frequent such distorted decisions are. The simulations have been run in Java and statistics carried out over many runs (typically 300,000) so that the standard errors of measured quantities are smaller than visible points in graphs, or than the last digit in tables.

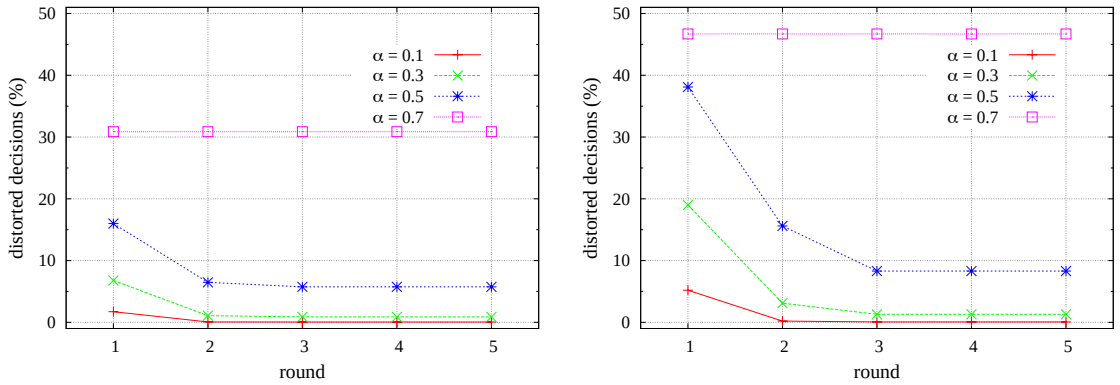
# 4 Holding a sufficient number of rounds

Suppose the members of the group speak once, one after another in a random order, typically because they are randomly arranged around a table. For non-random orders, cf. Section 5. Furthermore, consider here that the expression of views is limited to a single round because time is short, discussions have already taken place in the past, or the chair simply thinks that one roundtable is enough, "if people have had their say why ask again?" Then for a 5-group, 2% of decisions are distorted for $\alpha$ equal to 0.1, 7% for $\alpha$ equals 0.3, 16% for $\alpha$ equals 0.5 and up to 31% for for $\alpha$ equals 0.7. For a misrepresentation rate of 0.5, 13% of decisions are distorted for a 3-group, 16% for a 5-group, 19% for a 9-group and 22% for a 25-group. To help interpret these numbers, note that the percentage of distorted decisions cannot be larger than 50% on average by symmetry. For instance, 16% ($n = 5$, $\alpha = 0.5$) corresponds roughly to a third of the maximum possible distortion, and can be seen as quite important. (Remember that opinion dynamics is absent from the present model and could still amplify these effects, cf. Section 3.) Thus, when one roundtable is held, the misrepresentation of views can induce a sizable proportion of distorted decisions, even with a moderate misrepresentation rate.

Intuitively, holding several roundtables could make a difference, because the first agent may change his or her public view in a conciliatory way, or because initial non-dissenters may finally dissent from the majority because others have dissented after them. Mathematically, when new agents publicly dissent, the average $G_i$ of the last expressed opinions becomes closer to 0.5. Thus, the quantity $(1-\alpha)p_i+\alpha G_i$ may finally cross the 0.5 threshold for some values of $i$, and the corresponding agents become new dissenters.

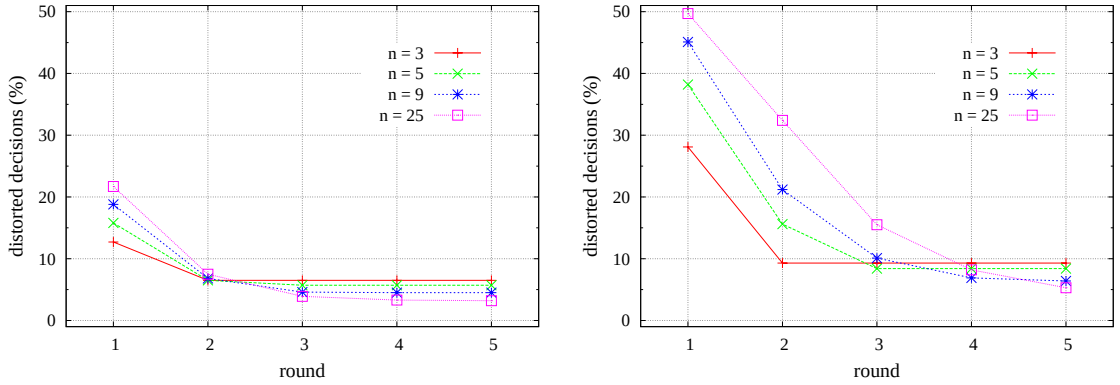
We analyzed how large this effect can be. Figure 1a quantifies it for a 5-group and various misrepresentation rates. Holding several roundtables helps: it decreases the percentage of distorted decisions, or at worst does not change it. This improvement can be substantial: for example, for $\alpha = 0.5$, the distortion is almost divided by 3 between round 1 and 3. Here, after 2 or 3 rounds, an equilibrium point is reached in which agents no longer change their public views. Some agents still misrepresent their private views, and no further improvement in the percentage of distorted decisions can be reached. The case of $\alpha = 0.7$ calls for specific comments. It can be shown that, for all $\alpha > 2/3$, the group's decision is merely the first speaker's private view[5] — hence the flat curve, with the same

[5]If $\alpha > 2/3$, once the first agent has expressed his or her view ($e_1 = \text{Proj}(p_1)$), whatever the private view $p_2$ of agent 2, the quantity $(1-\alpha)p_2+\alpha G_2$ is on the same side of 0.5 as $p_1$, and yields the same public




(a) Random order. (b) Increasing or decreasing order.

Figure 1: Influence of the number of rounds and of the misrepresentation rate on the percentage of distorted decisions ($n = 5$ agents).




(a) Random order. (b) Increasing or decreasing order.

Figure 2: Influence of the number of rounds and of the size of the groups on the percentage of distorted decisions ($\alpha = 0.5$).

value for all $\alpha > 2/3$. This happens for example when social influence is too high, the most influential agent speaks first, and all other agents follow him or her.

Consider next the effects of the size of the group for $\alpha = 0.5$ (Figure 2a). Groups like expert panels or committees are typically not too large, so we vary $n$ to as high as 25. Holding several rounds also decreases the percentage of distorted decisions by a factor 2 (for $n = 3$) to 7 (for $n = 25$) for the range of parameters studied here. After a few rounds of discussion, an equilibrium is reached again and additional rounds do not lead to further improvement beyond this point. A specificity of large groups is that distortion is significantly higher at the beginning but lower at equilibrium (in the Appendix, we state and prove several propositions about this general trend).

Overall, holding several rounds, typically 2 or 3, is a simple but significant way to

view.

reduce the effects of the misrepresentation of views, and insisting on many roundtables is pointless. This morale is robust for various sizes of groups and misrepresentation rates, below a misrepresentation threshold over which nothing can be done and all agents follow the first speaker. So, a chairman who does not want to make another roundtable after the first one has brought a near-consensus but may well be wrong. Holding several rounds can reduce misrepresentation effects within deliberations, show the near-consensus to be merely apparent, and reverse an oral distorted vote.

# 5 Speaking in a random order

So far in this paper, agents have been assumed to speak in a random order. In real groups, views and order of speech are often correlated. For instance, when sitting around a table, people who feel close may sit next to each other; or the chair may let those who hold similar views speak first, possibly to create a momentum in the deliberation procedure. To study the impact of such correlations on the magnitude of misrepresentation effects, we simulated the model when agents are positioned according to the two symmetric cases of increasing or decreasing private views, Then, distortion can be expected to be maximum.

As before, we explored the parameter space for $n$ and $\alpha$. Figure 1b shows the evolution of distortion for a 5-group and different misrepresentation rates, to be compared to Figure 1a with a random order. When agents speak in an increasing or decreasing order, for $\alpha = 0.5$, the percentage of distorted decisions can be 2.5 times as high as with a random order (after round 1) and is typically 1.3 times as high after the transient regime. Thus, our suggestion to organize several roundtables holds all the more for non-random orders. When $\alpha$ is larger than 2/3, the first speaker fully determines the result (cf. Section 4) and the percentage of distorted decisions approaches 50%.[6]

Figure 2b shows that similar patterns occur for a fixed misrepresentation rate, when the size of the group varies. As with a random order, distortion is higher for increasing or decreasing orders. It is also higher for large groups for the first rounds, but becomes lower when the equilibrium regime is reached, which happens after one or two more rounds. Overall, for all orders, the stable regime is more virtuous for large groups but more costly to reach in terms of number of individual interventions and deliberation length. After 5 rounds, the distortion is 5.3% for a 25-group and 9.5% for a 3-group (in an increasing or decreasing order), but it takes 125 oral interventions in the former case and 15 in the latter.

Importantly, convergence towards equilibrium is also slower for increasing or decreasing orders and takes 3 or more rounds. This means that a random order reduces distortions and does so more quickly. For example, for a 9-group, after 2 rounds, the distortion is 5.7% for a random order but still 6.4% after 5 rounds for an increasing or decreasing order. So, globally, a random order is more efficient and can save deliberation time.

[6] The exact percentage is $1/2 - 1/2^n$. Proof: without loss of generality, consider the increasing order. Assuming an odd number of agents, there are two cases. First case: a majority of agents would express 0.25 without misrepresentation. This occurs with a probability of 1/2. With an increasing order, the first speaker expresses 0.25 and there is no distortion in the group's decision. Second case: there is a majority of 0.75 in the group. There is no distortion if and only if no speaker privately favors 0.25 because he or she would speak first. This happens with a probability of $1/2^n$. QED.

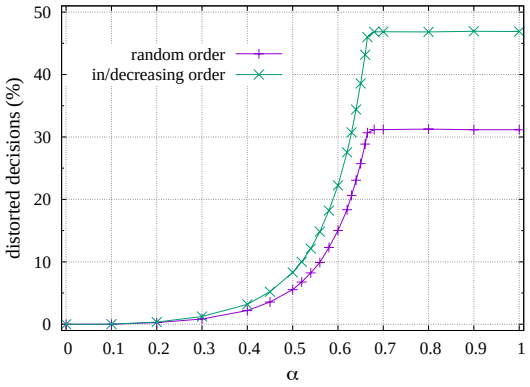



Figure 3: Distortion after the third round ($n = 5$). (Comparisons between the two orders should be made along vertical lines.)

Finally, it should be noted that the random order is also fairer. Indeed, cases of distortions are then accidental and equally in favor of 0.25 or 0.75. Thus, the average of decisions is 0.5, as it should be with a symmetric distribution of private views. By contrast, increasing or decreasing orders generate one-sided errors in favor either of 0.25 or 0.75. For example, 16% of errors with an increasing order means that 66% (that is 2/3) of decisions are in favor of 0.25 and 33% (that is 1/3) in favor of 0.75. With such orders, decisions are on average more distorted *and* one-sidedly biased, and the corresponding groups will seem to be clearly polarized (whereas the distribution of private views within them is not).

The increasing and decreasing orders correspond to extreme situations, and actual cases may lie somewhere between them and the random order. In any case, the larger the group, and the more likely manipulations or correlations in the order of speech, the more advisable it is for groups to adopt a procedure that guards them against distortion-inducing orders of speech. Our results show that a simple but efficient way to avoid correlations and manipulations, to reduce distortion and to make deliberations more efficient and fairer is to adopt a random order of speech, especially when deliberative time is short.

## 6 Making deliberation dissenter-friendly

To decrease distortion, one can also try to reduce the misrepresentation rate. As indicated in Section 2, misrepresentation can stem from various, potentially concurring factors such as a fear of retaliation, an unwillingness to be seen publicly as a dissenter, or a feeling that dissenting is not worth the effort. While some of these factors are hard to overcome, even when agents are aware of them, others can be partly acted upon. For example, we can try to make deliberations as unabrasive and non-conflictual as possible, create a friendly atmosphere, and explicitly and benevolently welcome dissenters and dissenting views so that agents feel the need to misrepresent their views less.

The questions that remain to be answered now are: to what extent can the unabrasive strategy be a key one, and is a small decrease in the misrepresentation rate worth the effort? It can be; as we now show, a small variation of $\alpha$ can make a big difference to the distortion.

Figure 3 shows the distortion rate for a 5-group after 3 rounds of speech, id est when distortion effects have roughly stabilized. For $\alpha > 2/3$, the distortion is maximal and stable.[7] For low values of $\alpha$, distortion is present but minor. The striking feature is that in the intermediate zone of approximately [0.4,0.7]), distortion does not rise linearly but becomes increasingly steep as $\alpha$ increases. This result is coherent with the finding that pluralistic ignorance and social behaviors rooted in misrepresentation effects can quickly disappear (Kuran, 1995). In this intermediate zone, a small change in $\alpha$ brings about a large change in the distortion rate. So, unless one is in extreme deliberative contexts, when agents barely or almost completely misrepresent their views, minor differences in misrepresentation can have large effects. Thus, trying to reduce even slightly the misrepresentation rate by simple methods, e.g. promoting adequate rules of deliberation demeanor, can be worthwhile, all the more so since one usually does not know where a deliberative group stands on the misrepresentation spectrum.

# 7 Expressing fine-grained views

So far we have assumed that agents simply express the option they favor, namely 0.25 or 0.75 (see Section 3.1). But agents have richer private views in $[0, 1]$. Would the group's decisions be improved and would the distortion rate be reduced if agents expressed finer-grained views? With 4 possible expressions, this amounts to being able to say "rather X" or "absolutely X", instead of just "X". Because this possibility reduces the discrepancy between agents' private and public views, one may think that it should also reduce the distortion rate and be epistemically beneficial. However, things are not so simple. We highlight below that it is not clear whether this intuitively appealing strategy is a genuine solution.

We considered 3 extensions of the model, in which agents can express a view among $v = 4, 8$ views or any view within $[0, 1]$. For $v = 4$ or 8, here again, the views are chosen so that they stand in the middle of $v$ equal-sized intervals, e.g. with $v = 4$, the views are $\{0.125, 0.375, 0.625, 0.875\}$. Agent $i$'s expressed view is still given by Equation 2, but the projection is now made on the 4 or 8 views, or there is no projection in the last case. The group's final decisions are still taken by the majority of agents binary votes, interpreting an expressed view above (resp. below) 0.5 as a vote for 0.75 (resp. 0.25).

The patterns discussed in Sections 4 to 6 are also found in these versions of the model. Two additional major points stand out. First, when the distortion rate is high, i.e. typically when $\alpha > 0.5$ (cf. Figure 3), allowing for finer-grained expression substantially lowers the distortion rate after a few rounds. For instance, for $\alpha = 0.7$, $n = 5$ and a random order, the distortion rate is 31% after 3 rounds with $v = 2$ possible expressions, but respectively 10, 10 and 12% for $v = 4, 8$ or infinity. What happens is that, when the misrepresentation rate is high, refined public views allow for easy and moderate dissent, damping the influence of the first speakers and preventing huge distortions from taking place systematically. Second, when the misrepresentation rate is moderate or low, typically when $\alpha \leq 0.5$, allowing for finer-grained expression slightly or moderately increases distortion. For instance, with $\alpha = 0.3$, $n = 5$ and a random order, distortion is 0.9% after 3 rounds for $v = 2$, and

[7] Cf. Section 4, footnote 5, and Section 5, footnote 6.

between 3 and 4% for $v = 4, 8$, or infinity.

These results call for conditional and cautious recommendations. A finer-grained expression of views may be seen as *per se* globally beneficial since it very slightly increases small distortion rates and significantly reduces large ones. It seems particularly appropriate when one suspects the misrepresentation rate $\alpha$ to be high. However, enabling a finer-grained expression of views opens the door to strategical voting, id est deliberately expressing extreme views to pull the group in one's direction; thus, in the absence of combined investigations of the two mechanisms, it cannot be flatly recommended in general. A more refined case-by-case analysis may be needed here to reach well-entrenched conclusions.

# 8 Discussion

We will now discuss further the robustness and plausibility of the results presented above. A potential issue is the precise value of the misrepresentation rate $\alpha$ and our decision to keep it constant. How much majorities and minorities cross-influence each other is still a debated question within social psychology, and it seems to call for a contextual answer. Social influence appears to depend on various factors such as the type of issue that is debated, the type of task, the degree of retaliation that dissenters can expect, or the type of agents involved (Bond, 2005). This is the reason why we have studied distortion for different misrepresentation rates.

Some models make influence depend on the size of the majority (Latané, 1981, Tanford and Penrose, 1984, Mullen, 1983), but in these models the growth of social influence on agents is very slow for groups larger than 3. In experimental cases like those investigated by Asch, the influence of the majority reaches a full impact as soon as $k = 3$. Our model is coherent with this feature, as it limits the influence of a majority to the fixed value of the misrepresentation rate $\alpha$. However, in Asch's experiments, for majorities smaller than 3, the social influence increases from 0 to the threshold value (Asch, 1951). How steep this transient increase is, and why, are still debated issues[8].

Accordingly, we also analyzed the case of a linear transient increase of the misrepresentation rate: during the first roundtable, the second and third speakers misrepresent their views with rate $\alpha/3$ and $2\alpha/3$ respectively, and then the misrepresentation rate equals $\alpha$. Our simulations show that this change reduces somewhat the distortion effect, which is understandable because the transient effect makes it easier for the first speakers to dissent. This drop in the percentage of distorted decisions can be significant for very small groups and a random order, e.g. for a 5-group and $\alpha = 0.5$, we note a reduction of 9% after 1 round, but this is temporary and transient, falling to 3% after 3 rounds, under the same conditions. Furthermore, the larger the group, the smaller this difference is, e.g. for a 25-group and $\alpha = 0.5$, the reduction is 5.5% after 1 round and 0.5% after 3 rounds. Importantly, for moderate values of $\alpha$, the progressive misrepresentation rate has very little impact on distortion for increasing or decreasing orders of speech, especially for large groups. For a 5-group and $\alpha = 0.5$, it amounts to 1.7% after 2 rounds, and 1.2% after 3 rounds. The explanation is that, in these situations, it is quite frequent for the first 3

[8] An important aspect seems to be whether the influence is of an informational type, where people comply not to be wrong, or of a normative type, where they yield to normative pressures.

speakers to express the same view, hence the transient regime makes no difference. Overall, the existence of this short-lived regime has mostly significant effects for small groups, when $\alpha$ is in the critical zone where distortion is maximal ($\alpha > 2/3$ for $n = 5$, see Figure 3) but $\alpha/3$ and $2\alpha/3$ are not. Be this as it may, even with this transient mechanism, simulations show that our morals remain: it is preferable to make at least 2 or 3 roundtables, to adopt a random order of speech, and to work in a dissenter-friendly atmosphere.

Finally, we consider the experimentally documented fact that, when one or several agents have already dissented publicly, other agents dissent more easily and conformity can significantly decrease (Asch, 1955), even if this effect is somewhat contextual[9]. In our model, the fixed misrepresentation rate corresponds to the weight given to the group's view, which is an average over expressed views. Accordingly, when dissenters emerge, the weight that is given by individuals to the public views of those with whom they disagree linearly decreases. Then, for a 5-group and $\alpha = 0.5$ (resp. 0.4; 0.3), if the first four speakers are unanimous, the probability that a fifth, while privately being a disagreeing speaker, misrepresents her view is 50% (resp. 33.3%; 21.4%). This percentage drops to 25% (resp. 16.7 %; 10.7%) if the fourth speaker has dissented. Thus, for such cases, the social support brought about by a previous dissenter divides the misrepresentation effect by 2. Overall, while the model is not aimed to reproduce the subtle and contextual variations of social influence with the degree of social support, it plausibly accommodates such a type of effect.

# 9 Conclusion

We have analyzed how much the conformist mechanism of view misrepresentation can warp deliberations and distort the decisions that are finally taken by groups. The model we introduced was not directed to make precise predictions in specific situations, but to identify general features of misrepresentation and suggest deliberative procedures to reduce the ensuing distortion in decisions. Our investigation feeds the conclusion that the effects of misrepresentation can be large. This provides more evidence for the claim that a consensual group decision can be warped if it is not issued from appropriate procedures that guarantee that minorities are heard (Beatty and Moore, 2010). Our results also point out procedures that decrease the harmful effects of view misrepresentation. If one wants to dampen distortion, it is advisable to make at least 2 or 3 roundtables, to speak in a random order, to make discussions inclusive and dissenter-friendly, and perhaps to find ways of letting agents express fine-grained views. Because of their robustness, these general suggestions do not depend on the exactness of a mapping between specific values of parameters and specific situations. Further, since agents do not usually have precise clues concerning how much other deliberators misrepresent their views or manipulate the order of speech, it is difficult for them to know where exactly they lie on the "misrepresentation and distortion" map. Thus, it may be a safe move to use these procedures by default, all the more since they are neither costly nor difficult to implement.

Importantly, making absolute normative claims about deliberation procedures usually

[9]The drop in the conformity effect when unanimity in the group is disrupted seems to be specifically important in Asch-like cases, when the tasks is a trivial one and is about facts, arguably because people do not expect opinions concerning the physical world to be different (Allen and Levine, 1968).

requires a global perspective. One needs to consider whether these procedures agree with rights, norms, or principles that we otherwise endorse. One also needs to take into account the global effects of these procedures on information exchanges, not just on misrepresentation and distortion. Hence, the above suggestions are merely *ceteris paribus* recommendations: to the extent that they do not have other detrimental effects, these procedures seem beneficial. Having said this, it is not clear how the suggested procedures of making people speak in a random order, organizing several rounds of expression of views, trying to limit misrepresentation of views, and perhaps encouraging the fine-grained expression of views, could significantly alter the quality of deliberations, quite the contrary. In the absence of additional evidence against them, considering that they are worth adopting seems reasonable. In any case, testing the effects of these procedures for deliberative groups of various types through empirical investigations would be welcome and would enable the above theoretical insights to be cross-checked. Investigating theoretically and experimentally what happens when the misrepresentation mechanism additionally feeds the dynamics of private views would also provide additional evidence about the global epistemic impact of misrepresentation on deliberative exchanges.

Finally, our results also suggest that misrepresentation can generate purely apparent consensus. Therefore, shortcutting the final voting stage is epistemically slippery, and asking for a secret ballot is often not superfluous. Nevertheless, as emphasized above, there can be strong reasons against secret voting. A synchronous public vote does not necessarily make social influence disappear. In a hand-raising vote, which is not fully synchronous, agents can follow other influential agents or just see what other agents do, like in standing ovation events. This can be grounds for adopting a two steps procedure in which agents vote secretly and individually disclose and explain their votes afterward to preserve accountability and publicity. This is the type of reform that was adopted in 2007 for FDA advisory committees[10]. Still, social influence and misrepresentation effects can take place both in the deliberation preceding the vote, and in the vote itself if the ballots are eventually disclosed and agents already have beliefs about what the others think. For this reason, applying the suggestions made above, or similar ones, to the deliberation that takes place before the vote can still be recommended.

[10]See again (Urfalino and Costa, 2015) for a detailed analysis of this reform and its effects.

# 10 Appendix

In Section 4, we noted that for large groups distortion is higher in the beginning and lower after several rounds of contributions, when equilibrium is reached. Here we state some general propositions concerning this trend.

**Proposition 1.** If agents speak in an increasing or decreasing order of private views, the percentage of distorted decisions at the end of the first roundtable tends to 50% when the size $n$ of the group tends to infinity.

Proof. First consider the increasing order. When $n$ tends to infinity, the probability that the first speaker says "0.25", and thus initializes the group's view to 0.25, tends to 1. Subsequent agents in the first roundtable will keep saying "0.25" as long as their private view is below the threshold value $t = (0.5 - 0.25\alpha)/(1 - \alpha)$; for $\alpha = 0.5$, $t = 0.75$. Because in the model the private views are picked from a uniform distribution between 0 and 1, as $n$ tends to infinity, the proportion of agents whose private views are below $t$ tends to $t$, with $t > 0.5$, and the probability that a majority of people have a private view below $t$ and say "0.25" tends to 1. In 50% of the cases, this will not coincide with a private vote, and will give distortion. The case of a decreasing order is symmetric.

**Proposition 2.** If agents speak in an increasing or decreasing order of private views, when the size $n$ of the group tends to infinity the percentage of distorted decisions decreases with the number of rounds $k$ and tends to 0 when $k$ tends to infinity.

Proof. With an increasing order, as noted above, agents keep saying "0.25" across the first roundtable (and the public view of the group remains 0.25) until their private views are above $t$. Then, agents say "0.75" and the public view of the group increases. During the second roundtable #2, the threshold value $t_2$ is lower than $t$, so more people say "0.75", and the public view of the group increases again. Things continue in this way for subsequent rounds. For the remainder of the proof, we assume $\alpha = 0.5$ for simplicity. When the size of the group tends to infinity, it can be shown easily that $t_{k+1} = 0.25 + 0.5t_k$. Therefore, the sequence $t_k$ decreases, tends to the fixed point 0.5, and distortion vanishes. What makes the thing work is that when $n$ tends to infinity, there are always new people whose private views lie between $t_k$ and $t_{k+1}$. These people stop misrepresenting their views and the dampening process never gets stuck. In contrast, the smaller the size of the group, the more often the process gets stuck and distortion remains considerable.

## References


Allen, Vernon L. and John M. Levine, 1968, "Social Support, Dissent and Conformity", *Sociometry* 31, 2, 138–49.

Asch, Solomon E., 1951, "Effects of Group Pressure on the Modification and Distortion of Judgments". In H. Guetzkow (Ed.), *Groups, Leadership and Men* (pp. 177–190), Pittsburgh, PA: Carnegie Press.

Asch, Solomon .E., 1955, "Opinions and Social Pressure", *Scientific American*, 193 (5): 31–35.

Banerjee, Abhijit V., 1992, "A Simple Model of Herd Behavior", *The Quarterly Journal of Economics*, 107(3), 797–817.

Beatty, John, 2006, "Masking Disagreement among Experts", *Episteme*, 3 (1-2), 52–67.

Beatty, John and Alfred Moore, 2010, "Should We Aim for Consensus?", *Episteme*, 7 (3), 198–214.

Bond, Rod, 2005, "Group Size and Conformity", *Group Processes & Intergroup Relations*, 8 (4), 331–354.

Bright, Liam Kofi, 2017, "On Fraud", *Philosophical Studies*, 174 (2), 291–310.

Cialdini, Robert. B., and Griskevicius, Vlad 2010, "Social influence", In *Advanced Social Psychology: The State of the Science*, edited by R. Baumeister and E. Finkel, 385–418, New York: Oxford Univ. Press.

Cohen, Joshua, 1986, "An Epistemic Conception of Democracy", *Ethics*, 97 (1), 26–38.

Deffuant, Guillaume, David Neau, Frédéric Amblard, et Gérard Weisbuch, 2000, "Mixing Beliefs among Interacting Agents", *Advances in Complex Systems*, 3 (1): 87–98.

Douven, Igor, 2009, "Introduction: Computer Simulations in Social Epistemology", *Episteme*, 6 (2), 107–109

Elster, Jon, 2015, *Secrecy and Publicity in Votes and Debates*, Cambridge University Press.

Estlund, David M., 2008, *Democratic Authority: a Philosophical Framework*, Princeton (N.J.): Princeton University Press.

Gilbert, Margaret, 1987, "Modeling Collective Belief", *Synthese*, 73 (1), 185–204.

Goldman, Alvin I., 1999, *Knowledge in a Social World*, Clarendon Press.

Hegselmann, Rainer and Krause, Ulrich, (2002), "Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation", *JASS*, 5 (3).

Hogg, Michael A., 2010, "Influence and leadership" In *The Handbook of Social Psychology.* 5th ed, vol. 2, edited by S. T. Fiske, D. T. Gilbert, and G. Lindzey, 1166–1206. Hoboken, NJ: Wiley.

Kosolosky, Laszlo, and Jeroen Van Bouwel, 2014, "Explicating Ways of Consensus-Making in Science and Society: Distinguishing the Academic, the Interface and the Meta-Consensus", in *Experts and Consensus in Social Science*, New York: Springer, eds: Carlo Martini and Marcel Boumans.

Kuhn, Thomas Samuel, 1962, *The Structure of Scientific Revolutions*, Chicago, the University of Chicago Press.

Kuran, Timur, 1995, *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Cambridge (Mass.), Harvard University Press.

Kuran, Timur, and Sunstein, Cass R., 1998, "Availability Cascades and Risk Regulation". SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Lackey, Jennifer, ed., 2014, *Essays in Collective Epistemology*, Oxford: Oxford University Press.

Landemore, Hélène, 2013, *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*, Princeton, N.J.: Princeton University Press.

Latané, Bibb, 1981, "Psychology of Social Impact", *American Psychologist*, 36, 343–356.

Lehrer, Keith and Carl Wagner, 1981, *Rational Consensus in Science and Society: a Philosophical and Mathematical Study*, Dordrecht.

List, Christian and Goodin, Robert E., 2001, "Epistemic Democracy: Generalizing the Condorcet Jury Theorem", *Journal of Political Philosophy*, 9 (3), 277–306.

List, Christian, and Pettit, Philip, 2011, *Group agency: the Possibility, Design, and Status of Corporate Agents*, Oxford: Oxford University Press.

Luskin, Robert C., Fishkin, James S., Hahn, Kyu S., 2007, "Consensus and Polarization in Small Group Deliberations", Prepared for presentation at the annual meeting of the American Political Science Association, Chicago, IL, August 30-September 2.

Martini, Carlo and Boumans, Marcel (eds), 2014, *Experts and Consensus in Social Science*, New York: Springer.

Mayo-Wilson, Conor, 2014, "Reliability of Testimonial Norms in Scientific Communities", *Synthese*, 191 (1), 55-78.

Mill, John Stuart, 1859, *On Liberty*, London, John W. Parker and Son, West Strand.

Muldoon, Ryan, 2013, "Diversity and The Division of Cognitive Labor", *Philosophy Compass*, 8(2), 117–125

Mullen, Brian, 1983, "Operationalizing the Effect of the Group on the Individual: a Self-Attention Perspective", *Journal of Experimental Social Psychology*, 19, 295-322.

Olsson, Erik, J., 2011, "A Simulation Approach to Veritistic Social Epistemology", *Episteme*, 8(2), 127–143

Schlesinger, Arthur M Jr., 2002, *A Thousand Days: John F. Kennedy in the White House*, "3. Cuba in the Cabinet Room", Boston: Houghton Mifflin.

Sunstein, Cass R., 2000, "Deliberative Trouble? Why Groups Go to Extremes", *The Yale Law Journal*, 110(1):71–119.

Sunstein, Cass R., 2005, *Why Societies Need Dissent*, Harvard University Press, chap. 8, "Are Judges Conformists Too?".

Sunstein, Cass R., 2010, "Deliberating Groups Versus Piction Markets (or Hayek's Challenge to Habermas)". In *Episteme*, edited by Alvin I. Goldman and Dennis Whitcomb, 3:192–213. 3. Oxford University Press.

Tanford, Sarah and Penrose, Steven, 1984, "Social Influence Model: A Formal Integration of Research on Majority and Minority Influence Processes", *Psychological Bulletin*, 95, 176–194.

Urfalino, Philippe and Costa, Pascaline, 2015, "Secret-Public Voting in FDA Advisory Committees," in *Secrecy and Publicity in Votes and Debates*, ed. Jon Elster, Cambridge University Press.

Weisbuch, Gérard, Deffuant, Guillaume and Amblard, Frédéric, 2005, "Persuasion Dynamics" *Physica A*, 353, 555–75.

Zollman, Kevin, 2010a, "Social Structure and the Effects of Conformity", *Synthese*, 172 (3), 317–340

Zollman, Kevin, 2010b, "The Epistemic Benefit of Transient Diversity", *Erkenntnis*, 72, 17–35

Zollman, Kevin, 2012, "Social Network Structure and the Achievement of Consensus", *Politics, Philosophy & Economics* 11 (1), 26–44.