

Anchoring in Deliberations

Stephan Hartmann* Soroush Rafiee Rad†

May 11, 2017

Abstract

Deliberation is a standard procedure to make decisions in not too large groups. It has the advantage that the group members can learn from each other and that, at the end, often a consensus emerges that everybody endorses. But a deliberation procedure also has a number of disadvantages. E.g., what consensus is reached usually depends on the order in which the different group members speak. More specifically, the group member who speaks first often has an unproportionally high impact on the final decision: She *anchors* the deliberation process. While the anchoring effect undoubtedly appears in real deliberating groups, we ask whether it also appears in groups whose members are truth-seeking and rational in the sense that they take the information provided by their fellow group members properly into account by updating their beliefs according to plausible rules. To answer this question and to make some progress towards explaining the anchoring effect, a formal model is constructed and analyzed. Using this model, we study the anchoring effect in homogenous groups (i.e. groups whose members consider each other as equally reliable), for which we provide analytical results, and in inhomogeneous groups, for which we provide simulation results.

1 Introduction

There are numerous instances of group decision-making in our daily practice. For example, families have to decide on where to spend the summer holidays, funding agencies have to decide which research projects to support, and juries in courts have to decide whether a defendant is guilty or not. Sometimes a decision is made in the light of different preferences (and each group member wants to get the best out of it for herself), and sometimes all group members share the conviction that the resulting decision should be best in some sense that is commonly agreed upon. Juries in court, for example, aim at making the *right* decision: Every jury member wants that a guilty defendant is convicted,

*Munich Center for Mathematical Philosophy, LMU Munich, Geschwister-Scholl-Platz 1, 80539 Munich (Germany) – <http://www.stephanhartmann.org> – S.Hartmann@lmu.de.

†Munich Center for Mathematical Philosophy, LMU Munich, Geschwister-Scholl-Platz 1, 80539 Munich (Germany) – <http://www.rafieerad.org> – Soroush.Rafiee.Rad@lrz.uni-muenchen.de.

and no one wants that an innocent defendant is sent to prison. Likewise, committees such as the Intergovernmental Panel on Climate Change (IPCC) deliberate to arrive at the best far-reaching policy recommendations (cf. Oreskes (2004)). Many real-life decisions are somewhere in between these extremes: Personal preferences play a role, but there is often also some overarching goal that all group members share. Decisions of funding agencies are a case in point. Here one wants, on the one hand, to fund the best possible research and the best possible researchers. On the other hand, some research topics are more important than others, some research institutes are better suited than others, and the corresponding judgements may depend on personal biases.

But how shall a group make a decision? Which decision-making procedure is best? Consider a group whose members have conflicting preferences. For example, after discussing the issue for some time, three family members want to spend the summer holidays on the sea, while one family member wants to spend some nice days in the mountains. However, all family members agree that they want to spend the holidays with each other. At this point it is unlikely that one of them will change her mind. And so the best procedure to arrive at a collective decision is to vote and decide what the majority prefers.

In other cases, a deliberation procedure will be better. Let us assume that all group members agree on a common goal and that additional personal preferences do not play a role. If, for example, all members of a jury in court have the goal to make the right decision, then each of them will appreciate the opportunity to learn from the other jury members. Proceeding in this way, some jury members will rationally change their minds as they may, e.g., become aware of new pieces of evidence or identify mistakes in the assumptions or in the reasoning they are making in the course of the deliberation. As a result, the group may eventually arrive at a better decision compared to the result of a straight-forward majority voting procedure. Sure, it will take longer to arrive at a decision (compared to voting) and a deliberation procedure is harder to implement (esp. for larger groups), but the possibility that the group members can change their beliefs in the light of what they learn from other group members has important advantages. What is more, often such deliberation procedures lead to a consensus that makes all group members happy as each of them endorses the decision of the group (Steele et al., 2007). The consensual decision emerges when the process of individual belief-changes comes to an end and all beliefs stabilize and coincide on a certain collective belief.

The optimal working of the deliberation procedure presupposes that all group members have one common goal (such as making the right decision) and that they make good use of the information provided by the other group members. As noted above, these conditions may not always hold in real deliberations. For example, some group members may have a hidden agenda that prevents them from saying what they consider to be true. Or a group member may not be able to properly take the information she receives from another group member into account because she gives the new information a much too high weight (and therefore effectively assigns a too high reliability to the information source).

Biases of this kind are hard to avoid. Some of them result from the fact that the individual group members are not fully rational as they are unable, e.g., to assess the

true reliability of the other group members. Other biases may result from the specific deliberation procedure that is adopted. Such biases may also show up in groups whose members are individually rational. Here we will focus on one such bias: the anchoring effect.

The anchoring effect (see, e.g., Bazerman (2002); Kahneman et. al. (2006); Tversky & Kahneman (1974)) is a widely studied effect in the literature on judgment and decision-making. For example, when asked to estimate at what age Mahatma Gandhi died, participants who were first asked whether Gandhi died before or after the age of 140 on average estimated that his age of death was 66.7. At the same time, participants who were first asked whether he died before or after the age of nine, on average estimated that his age of death was 50. The effect even shows up when a totally unrelated number is mentioned to the participants. A recent review, discussing this and many other examples, concluded that “anchoring effects are among the most robust and ubiquitous psychological phenomena in judgement and decision making” (Mussweiler et al. (2004, p. 196)). The effect shows up in many different contexts and many different phenomena can be described as instances of an anchoring effect. But what explains the effect? Mussweiler et al. (2004, p. 187) elaborate:

It is important to note, however, that these phenomena are not sufficiently explained by evoking an unspecific notion of anchoring. As such, the anchoring notion does not illuminate the underlying mechanisms, but only describes the direction of the observed influence (assimilation). In this respect, the term “anchoring” constitutes a descriptive rather than an explanatory concept which does not go beyond the terms assimilation and contrast (...). In order to be used as an explanatory concept, however, the psychological mechanisms that underlie anchoring first have to be sufficiently understood.

This article aims at making some progress on explaining a specific instance of the anchoring effect by constructing a *formal model*. More specifically we are interested in the anchoring effect in deliberations. This is the effect that typically, in a deliberation procedure the group member who speaks first has the highest impact on the resulting group decision:¹ She anchors the deliberation and the other group members refer to the judgment of the first group member and take it into account when they speak. We ask: does this effect also show up in a group of individually rational agents? To address this question, empirical studies do not suffice as the participants in experiments cannot be assumed to be individually rational. We have to construct a model that models the group members *counterfactually* as rational agents and that explores the consequences of this assumption with the help of mathematical analysis. Studying our model will help understanding the anchoring effect better and shed light on the psychological mechanisms that drive the effect. Our strategy is supported by arguments from the philosophy of science to the effect that counterfactual considerations play an important role in scientific

¹Arguably, also the last speaker has an important impact on the final decision. See, e.g. Mussweiler et al. (2004, p. 183). As we will see below, understanding this instance of the anchoring effect requires a model that differs from the model that we present in this article.

explanation and understanding (Frigg & Hartmann, 2006; Reutlinger, 2016; Reutlinger et al., 2017).

The remainder of this article is organized as follows. Section 2 discusses two models of deliberation on the basis of which our own model will be built. Section 3 presents our deliberation model which will then, in Section 4, be used to study the anchoring effect. Section 4.1 presents analytical results for the anchoring effect in homogenous groups (i.e. in groups whose members consider each other as epistemic peers) and Section 4.2 presents simulation results for inhomogeneous groups. Finally, Section 5 summarizes the main results of this article and points out some open problems and questions which should be addressed in future research.

2 Models of Deliberation

In this section, we consider two models of deliberation in more detail. It turns out that none of them can be used directly to study the anchoring effect in deliberations. However, we will see that the model we propose in the next section combines ideas of both models discussed now.

We begin with the *Lehrer-Wagner model of deliberation* (Lehrer, 1976; Lehrer & Wagner, 1981). Consider a group of n members, ordered from 1 to n , that has to fix the value of a real-valued parameter x . Initially, each group member i submits an *initial value* $x_i^{(0)}$ (with $i = 1, \dots, n$). Each group member i also assigns a normalized *weight* w_{ij} to each group member j (including herself). These weights sum up to one, i.e.

$$\sum_{j=1}^n w_{ij} = 1.$$

The weights reflect to what extent group member i takes the opinion of group member j into account. One can also think of them as group member i 's subjective estimates of the suitably normalized reliabilities of all group members. Now, each group member i updates her initial value $x_i^{(0)}$ in the following way (“linear pooling”) to obtain the new value $x_i^{(1)}$:

$$x_i^{(1)} = \sum_{j=1}^n w_{ij} x_j^{(0)} \quad (1)$$

If we denote the vector of initial values by $\vec{V}^{(0)} := (x_1^{(0)}, \dots, x_n^{(0)})^T$, the vector with the updated values by $\vec{V}^{(1)} := (x_1^{(1)}, \dots, x_n^{(1)})^T$, and the weight matrix by $(\mathbf{W})_{ij} := w_{ij}$, then we can write eq. (1) in a convenient matrix form:

$$\vec{V}^{(1)} = \mathbf{W} \vec{V}^{(0)} \quad (2)$$

Next, we iterate this process and obtain for deliberation round k :

$$\begin{aligned} \vec{V}^{(k)} &= \mathbf{W} \vec{V}^{(k-1)} \\ &= \mathbf{W}^k \vec{V}^{(0)} \end{aligned} \quad (3)$$

Lehrer and Wagner have shown that this process converges if the matrix \mathbf{W} is connected; see Lehrer & Wagner (1981). In this case, all entries in $\lim_{k \rightarrow \infty} \vec{V}^{(k)}$ are identical and all group members agree on this consensual value. Here are three remarks about the Lehrer-Wagner model:

1. In each round, all group members update at the same time. Hence, the anchoring effect cannot be studied in this model.
2. The weight matrix \mathbf{W} is not derived. It simply contains the weights each group member assigns to the members of the group. This is disturbing as one might expect that the weights a rational group member assigns to all group members are based on an estimate of their true reliabilities.
3. Lehrer and Wagner considered the model to be a model of rational consensus formation. Disagreement is impossible if the weight matrix is connected and the various group members do not assign, e.g., a weight of 1 to themselves. This requirement reflects the assumption that each group member is willing to learn from the other group members. As a result, dissent is impossible in a group of rational agents, which is too strong too strong, e.g. in situations of high uncertainty.

There are alternative models for the determination of a real-valued parameter by a group, including the Hegselmann-Krause model (or “bounded confidence model”); see Hegselmann & Krause (2002, 2006, 2009). This model has a number of interesting features. However, it uses a synchronic updating rule similar to eq. (3) and can therefore also not be used to study the anchoring effect.

Let us now turn to a model of deliberation presented in Hartmann & Rafiee Rad (2017). This model considers a group of n members, ordered from 1 to n , who deliberate in a sequence of rounds to eventually agree upon the truth value of a certain propositional variable (e.g. guilty/not guilty). Each group member enters the deliberation process with a prior probability of the proposition in question and then votes for or against the proposition based on that probability. It is also assumed that each group member has a certain *first order reliability* $r_i \in (0, 1)$ to make the right decision as well as a *second order reliability* $c_i^{(k)} \in (0, 1)$ which reflects how well group member i is able to estimate the first order reliability of the other group members $j \neq i$ in round k . Group members with a high value of $c_i^{(k)}$ give a more accurate assessment of the reliabilities r_j of the other group members than group members with a low value of $c_i^{(k)}$. More specifically, the model assumes that group member i estimates the reliability of a group member j in round k , denoted by $r_{ij}^{(k)}$, by drawing from a uniform distribution² around the true value r_j :

$$r_{ij}^{(k)} = \text{UniformDistribution} \left[\max(0, r_j + c_i^{(k)} - 1), \min(1, r_j - c_i^{(k)} + 1) \right] \quad (4)$$

²Here we deviate from the model in Hartmann & Rafiee Rad (2017) by taking the estimations from a uniform distribution instead of a β -distribution.

Thus, for $c_i^{(k)} = 0$, the uniform distribution extends over the whole interval $(0, 1)$. In this case group member i assigns a random reliability between 0 and 1 to group member j . Turning to the other extreme, for $c_i^{(k)} = 1$, we obtain $r_{ij}^{(k)} = r_j$, i.e. group member i gives an accurate assessment of the first order reliability of group member j . Taking this into account, each group member updates her prior probability by conditioning on the votes of the other group members. (Clearly, the better a group member is in estimating the reliabilities of the other group members, the better her updated probability will be.) This process is then iterated over several rounds, assuming that the second order reliabilities increase from round to round, reflecting the fact that one learns more and more about the reliability of the other group members in the course of deliberation, until a consensus is reached. If no consensus is reached after K_0 rounds, then a majority vote decides.

Hartmann and Rafiee Rad use this Bayesian model to investigate the question under which conditions a deliberation process tracks the truth better than a simple voting procedure that is supported by the Condorcet Jury Theorem. This question does not concern us here. We will also not discuss alternative Bayesian models of deliberation, such as the Laputa model (see, e.g., Olsson (2011)), as they are not relevant for the study of the anchoring effect. However, we will use the idea of first and second order reliabilities in our model of the anchoring effect to which we turn now.

3 Modeling Anchoring

Consider a group of n members, ordered from 1 to n , which has to fix the value of a real-valued parameter x . Each group member i enters the deliberation process with an *initial value* $x_i^{(0)}$. These values are not announced publicly right away. As in the Hartmann-Rafiee Rad model discussed in the previous section, each group member i is characterized by two reliabilities:

1. A first order reliability $r_i \in (0, 1)$ that measures how good group member i is in identifying the right value x .
2. A second order reliability $c_i^{(k)} \in (0, 1)$ that measures how good group member i is in round k in estimating the first order reliabilities of the other group members.

We present the details of our model in four steps.

1. **The Deliberation Procedure in a Nutshell.** The deliberation proceeds in a number of rounds. Each round consists of n steps. In round 1, step 1, group member 1 announces her assignment and provides reasons for it. Then all other group members update their assignments according to rational rules specified below, taking this new information into account. In step 2, group member 2 announces her assignment and provides reasons for it. Then all other group members update their assignments according to rational rules specified below, taking this new information into account. And so on until the end of round 1. Before round 2 starts, the

second order reliabilities $c_i^{(0)}$ are increased. This takes into account that the group members get to know each other better in the course of the deliberation (e.g. by assessing the reasons the group members give for their assignments) which has the effect that their ability to assess the first order reliabilities of the other group members improves.³ In fact, we assume that the second order reliabilities increase after every round until a maximum value $C_i \leq 1$ is reached after K rounds. Afterwards, $c_i^{(k)}$ remains constant. We use the following formula:⁴

$$c_i^{(k)} = \begin{cases} (C_i - c_i) \cdot k/K + c_i & : 0 \leq k \leq K \\ C_i & : k > K \end{cases} . \quad (5)$$

Note that updating the second order reliabilities justifies that the deliberation process proceeds in several rounds. In each round, the group members learn something new about the reliabilities of their fellow group members. And this is why they keep on updating. However, it seem natural to stop the updating procedure after a finite number $K_0 > K$ of rounds. Clearly, the value of K_0 will depend on contextual factors such as how patient the individual group members are, how much time there is for the deliberation, etc. If no consensus is reached after K_0 rounds, then the straight average of all assignments in the last round will be taken.

2. **Discrete First-Order Reliabilities.** While each group member i has a first order reliability $r_i \in (0, 1)$, we assume that the reliabilities that are actually used by the group members have discrete values. This simplifies the computations and is psychologically more plausible. More specifically, we assume that there are only three possible reliability values: H ('High'), M ('Medium'), and L ('Low'). We furthermore assume that every group member i has perfect access to her own first-order reliability⁵ and assigns herself a reliability of

- H if $r_i \geq 2/3$,
- L if $r_i \leq 1/3$, and
- M otherwise.

3. **Estimating the Discrete First-Order Reliabilities of Other Group Members.** Using the procedure described in the previous section, each group member i estimates the first order reliability $r_{ij}^{(k)}$ of the other agents in round k . This value is mapped onto H , M , or L as described above.

³Note that we do not model the process of giving reasons in detail. This would be the task of a more complex model. Instead, we focus on the *epistemic effect* of giving reasons, i.e. that the individual group members become better in assessing each others reliabilities. In this sense our model is *effective* or *phenomenological*. See Frigg & Hartmann (2006) for a discussion of the value of such models.

⁴It turns out that the results of our simulations, presented in Section 4.2, do not depend on the details of this formula. Other monotonically increasing functions of k lead to similar results. The linear dependence is chosen for reasons of simplicity.

⁵This is a strong assumption which can easily be relaxed. Indeed, psychological evidence suggests that people are not that good in assessing their own reliabilities. Notice, however, that we only need to assume that each group member has access to her own reliability bracket (i.e. H , M , or L) as opposed to the exact value of her own first order reliability since only those are used in the computations.

4. **The Updating Procedure.** After one group member announced her assignment, all other group members update their original assignments and their own reliability assignment according to the following rules. The rules express group member i 's assignment of the value of the parameter $x_i^{(k+1)}$ in round $k+1$ in terms of her assignment in round k and the assignment of the person who speaks (= group member j , a.k.a. the presenter) in round k . They are inspired by Elga's contribution to the debate about rational disagreement in the literature on social epistemology, see Elga (2007).

Presented in first person perspective, the updating proceeds as follows:

- (a) I am H (M , or L)⁶ and the presenter is my *epistemic peer*, i.e. she is (on my estimation) also H (M , or L). In this case, my new assignment is the straight average of her assignment and my original assignment:

$$x_i^{(k+1)} = \frac{1}{2} \left(x_i^{(k)} + x_j^{(k)} \right).$$

My reliability value (i.e. the reliability value that I use in the next round to update my assessment) remains H (M , or L).

- (b) I am H , the presenter is (on my estimation) L . In this case I disregard the opinion of the presenter and stick to my original assignment:

$$x_i^{(k+1)} = x_i^{(k)}.$$

My reliability remains H .

- (c) I am L , the presenter is (on my estimation) H . In this case I accept the opinion of the presenter:

$$x_i^{(k+1)} = x_j^{(k)}.$$

My reliability changes to H .

- (d) I am H (M), the presenter is (on my estimation) M (L). In this case, my new assignment is the weighted average of her and my original assignment:

$$x_i^{(k+1)} = \frac{1}{4} \left(3x_i^{(k)} + x_j^{(k)} \right).$$

My reliability value remains H (M).

- (e) I am L (M), the presenter is (on my estimation) M (H). In this case, my new assignment is the weighted average of her and my original assignment:

$$x_i^{(k+1)} = \frac{1}{4} \left(x_i^{(k)} + 3x_j^{(k)} \right).$$

My reliability value changes to M (H).

⁶For ease of writing, we assume that I (i.e. group member i) am one of the other group members, i.e. "I am H " is short hand for "My reliability value is H " etc.

Rule (a) expresses the Equal Weight View (Elga, 2007). Rules (b) and (c) are inspired by Elga’s discussion of the guru case. Rules (d) and (e) use weights that reflect that the reliability assignments in question are one step apart from each other.

Before we present our results, we would like to make two remarks:

1. At the end of round 1, we have effectively generated a Lehrer-Wagner weight matrix \mathbf{W} as the updating in each step can be represented as a matrix, and at the end of round 1 n such matrices were multiplied. Our model therefore provides a way to generate the weight matrix of the Lehrer-Wagner model.
2. Our model assumes discrete reliability values. This is psychologically plausible and simplifies the computations. At the same time, using discrete reliabilities disregards valuable information. We will show below that the anchoring effect is a consequence of our model, but one may wonder whether the effect disappears if one models more rational agents, i.e. agents who update on the basis of more fine grained reliability brackets or who even update directly on the basis of the (estimated) real-valued reliabilities. Perhaps the effect is, after all, only the result of the fact that the group members are not fully rational. We will revisit this question at the end of Section 4.2.

4 The Anchoring Effect in Deliberations

Let us now explore the consequences of our model. To do so, we study two types of groups separately: (i) In *homogeneous groups*, the different group members consider each other as epistemic peers. This case can be analyzed analytically and we will present several theorems in Section 4.1. (ii) In *inhomogeneous groups*, different group members have different reliabilities and we will present the results of our computer simulations in Section 4.2.

4.1 Homogeneous Groups

In a homogenous group, all group members are epistemic peers and only Rule (a) from the previous section applies. Let $\vec{V} := (x_1, \dots, x_n)^T$ be the vector of the assignments of the group members $1, \dots, n$ at some point in the deliberation process. Let us now assume that group member i speaks and all other group members update their assignments on that information. As all group members are epistemic peers, the new assignments are the straight average of the previous assignment and the previous assignment of the speaker. Denoting the new vector by $\vec{V}' := (x'_1, \dots, x'_n)^T$, this can be formally expressed as follows:

$$\vec{V}' = 1/2 B_n^i \vec{V}$$

Here $B_n^i := A_n^i + I_n$, where A_n^i is an $n \times n$ matrix with 1’s in the i^{th} column and zeros elsewhere. I_n is the $n \times n$ unit matrix.

Let us now denote the vector of initial assignments by $\vec{V}^{(0)} := (x_1^{(0)}, \dots, x_n^{(0)})^T$. The vector $\vec{V}^{(1)}$ after one round of deliberation is then given by

$$\begin{aligned}\vec{V}^{(1)} &= \frac{1}{2^n} (B_n^n \cdot B_n^{n-1} \dots B_n^1) \cdot \vec{V}^{(0)} \\ &=: \mathbf{B} \cdot \vec{V}^{(0)},\end{aligned}\tag{6}$$

with the weight matrix \mathbf{B} . Note the formal analogy to eq. (2). After k rounds of deliberation, we obtain⁷

$$\vec{V}^{(k)} = \mathbf{B}^k \cdot \vec{V}^{(0)}.\tag{7}$$

Building on the results of Lehrer and Wagner, we first show that the deliberation process converges.

Theorem 1 *The deliberation process described above converges to a consensus.*

Proof. It is easy to check by induction on k that $\prod_{i=1}^k B_n^i$ has exclusively positive entries in columns $1, \dots, k$. Hence \mathbf{B} only has positive entries. The theorem then follows from Theorem 7.1 in (Lehrer & Wagner, 1981, p. 130). ■

Next, we investigate the anchoring effect. To do so, we have to show that the first speaker gets a higher effective weight in the final weight matrix ($\lim_{k \rightarrow \infty} \mathbf{B}^k$) than all other speakers. This is stated in the following theorem.

Theorem 2 *In the consensus value reached by a homogeneous group through the process of deliberation described above, the opinion of the first speaker receives a higher weight than the opinion of any other group member. Moreover the opinion of the i^{th} speaker receives a higher weight than all subsequent speakers.*

The proof is given in Appendix A.

Finally, we show that Theorem 2 also holds for small deviations from the equal weight assignment. If the group members consider small deviations from the equal weights by assigning $(1 - \epsilon)/2$ weight to the speaker and $(1 + \epsilon)/2$ to themselves (for some small ϵ), then the above updating procedure can be represented as before. In this case, the update will be given by a matrix multiplication as before, replacing the matrices B_n^i by

$$B_n^{i'} := B_n^i - \epsilon E_n^i,$$

where

$$E_n^j := B_n^j - 2A_n^j.$$

⁷Note that the second order reliabilities do not change here. The group members consider each other as epistemic peers during the whole deliberation process. Our discussion is therefore analogous to the original Lehrer Wagner model where the weight matrix is the same in all deliberation rounds. This assumption is not uncontroversial, see, e.g. Hartmann et. al. (2009) and references cited therein. Lehrer & Wagner (1981) also relaxed the assumption and investigated under which conditions a consensus emerges nevertheless in their model. In the model considered in this section, the assumption allows us to derive some interesting analytical results. For results which do not rely on this assumption, see Section 4.2.

We then obtain after one round of deliberation:

$$\vec{V}^{(1)} = \frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \vec{V}^{(0)}$$

After k rounds of deliberation, the assignments are given by

$$\vec{V}^{(k)} = \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)}.$$

The following theorem states our result:

Theorem 3 *For a sufficiently small ϵ , the process of deliberation described above with weights set as $(1 - \epsilon)/2$ and $(1 + \epsilon)/2$ will converge to a consensus. That is if*

$$\vec{V} = \lim_{k \rightarrow \infty} \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)}$$

then $\langle \vec{V} \rangle_i = \langle \vec{V} \rangle_j$ for $i, j = 1, \dots, n$. Moreover in the consensus the initial assignment of each speaker receives a higher weight than the initial assignment of all subsequent speakers.

The proof and some more details are given in Appendix B.

4.2 Inhomogeneous Groups

Let us now turn to inhomogeneous groups where the different group members have different first order reliabilities. Here each group member has to estimate the first order reliabilities of the other group members to properly take their verdict into account. As it is not possible to obtain analytical solutions in such cases, we run computer simulations. The results will, of course, strongly depend on the distribution of reliabilities one starts with. We will study, for example, groups in which the first speaker has a lower first order reliability than the rest of the group as well as the groups in which some members are more reliable than the first speaker and some members which are less reliable than the first speaker. Of course, there will be reliability assignments for which the emergence or non-emergence of the anchoring effect is trivial. If the first speaker is highly reliable while the rest of the group has very low reliabilities then they will all simply adopt the assignment of the first speaker and the anchoring effect is inevitable. Similarly, the anchoring effect will not emerge if the first speaker has an extremely low reliability and the rest of the group members are all highly reliable. Thus the interesting cases are those groups with mixed reliabilities of high, medium and/or low in such a way that a considerable part of the group will not abandon their own assignments in favor of that of the first speaker. It is in particular interesting to see the emergence of the anchoring effect in groups where a large portion of the group members have reliabilities higher than the first speaker (but not high enough to completely discard her opinion).

Let us now present the results of our simulations of the deliberation process for different groups. We compute the probability that the final consensus of the group is closest to the initial assignment of the first speaker (call this proposition A) and plot this probability as a function of the group size unless stated otherwise. We shall denote the same proposition for the middle and the last speakers by A^{middle} and A^{last} respectively. The second order reliabilities are assumed to be the same for all group members. Notice that a second order reliability of 0.8 means that the estimation is made in intervals of length of at least 0.2. This means that the group members can possibly assign wrong reliability brackets to their fellow group members. The simulation algorithm and some more technical details are described in Appendix C.

Figure 1 shows the anchoring effect for a group in which the first speaker has a reliability of 0.85 and the rest of the group have reliabilities of 0.75 and 0.4, equally distributed. So at least half of the group members have reliabilities close to the reliability of the first speaker. Given the assignment of reliability brackets, the first speaker will on average be assigned high reliability (H), half of the group will be given reliabilities high or medium (H or M) and the other half will be given reliabilities M or L .

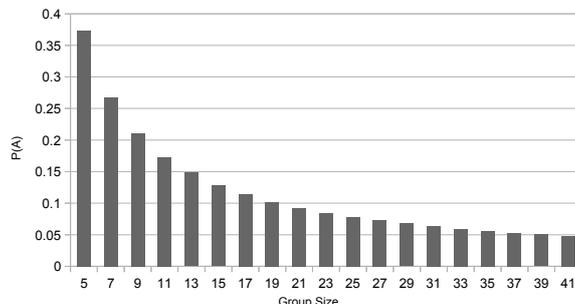


Figure 1: The probability of A as a function of the group size. The first speaker has a reliability of 0.85; the other group members have reliabilities of 0.75 and 0.4 (equally distributed).

Figures 2(i) and 2(ii) show the same group as in Figure 1, where the group member with high reliability (0.85) is the middle speaker and the last speaker respectively. The plots in Figure 2(i) and 2(ii) show the probability that the final consensus is closest to the original submission of this group member (the middle speaker and the last speaker respectively).

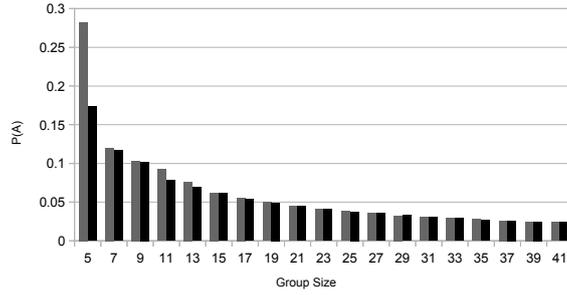


Figure 2: The probability of A^{middle} and A^{last} as a function of the group size. i) (lighter bars) The middle speaker has a reliability of 0.85; the other group members have reliabilities of 0.75 and 0.4 (equally distributed). ii) (darker bars) The last speaker has a reliability of 0.85; the other group members have reliabilities of 0.75 and 0.4 (equally distributed).

Comparing Figure 1 with Figures 2(i) and 2(ii) shows how the anchoring effect depends on the speaker's position in the group. In particular, the comparison of the plots suggests that the anchoring effect depends more on the speaker's position in the group than on her reliability as the same member placed in the middle or as the last speaker will result in a much smaller anchoring effect. To see this more clearly one can compare the anchoring effect for the first speaker in Figure 1 with that of the middle and the last speaker in Figures 3(i) and 3(ii) respectively, where the middle and the last speakers have significantly higher reliabilities than the rest of the group.

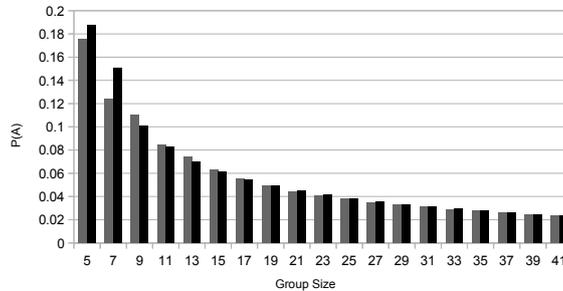


Figure 3: The probability of A^{middle} and A^{last} as a function of the group size. i) (lighter bars) The middle speaker has a reliability of 0.85; the other group members have reliabilities of 0.55 and 0.4 (equally distributed). ii) (darker bars) The last speaker has a reliability of 0.85; the other group members have reliabilities of 0.55 and 0.4 (equally distributed).

Next we study how the first speaker's reliability influences the anchoring effect. The first speaker in Figure 1 has a higher reliability than the rest of the group. This, however, is not necessary for the anchoring effect; the effect even emerges when the first speaker is

not particularly reliable in comparison with the other group members. Figure 4 shows the anchoring effect for two groups in which all (Figure 4(i)) or a considerable number of group members (Figure 4(ii)) have reliabilities higher than the first speaker.

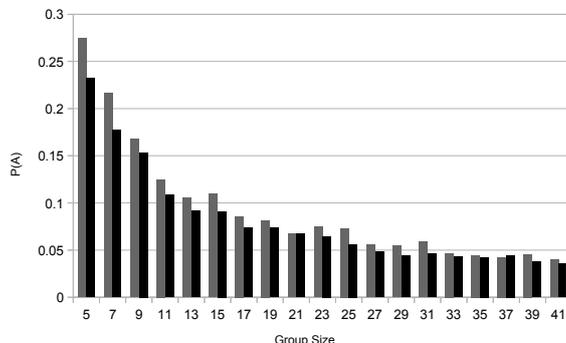


Figure 4: The probability of A as a function of the group size. i) (lighter bars) The first speaker has a reliability of 0.6; the other group members have reliabilities of 0.9. ii) (darker bars) The first speaker has a reliability of 0.5; other group members have reliabilities of 0.9 and 0.7 (equally distributed).

Although the anchoring effect also shows up when the first speaker does not have a particularly high reliability, the comparison of Figures 1 and Figure 4 suggests a positive correlation between the reliability of the first speaker and the intensity of the anchoring effect as one would expect. To see this more clearly, Figure 5 shows the probability of A as a function of the first speaker's reliability for a group of size 11 where the rest of the group members have reliabilities of 0.7 and 0.5.

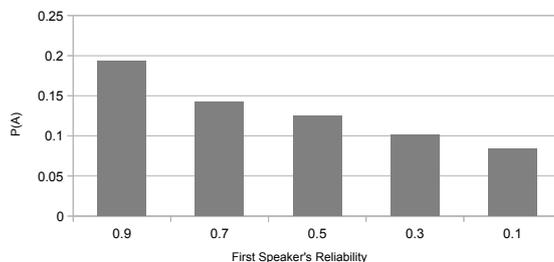


Figure 5: The probability of A as a function of the first speaker's reliability for a group of size 11. The other group members have reliabilities of 0.7 and 0.5 (equally distributed).

The next two graphs show the anchoring effect for two groups with random distributions of reliabilities. In Figure 6, the reliabilities of the group members are coming from a uniform distribution over the interval $(0, 1)$. In Figure 7, the reliabilities are assigned through a β -distribution with parameters $\alpha = \beta = 2$ over the interval $(0, 1)$. In both cases the first speaker has the reliability of 0.5.

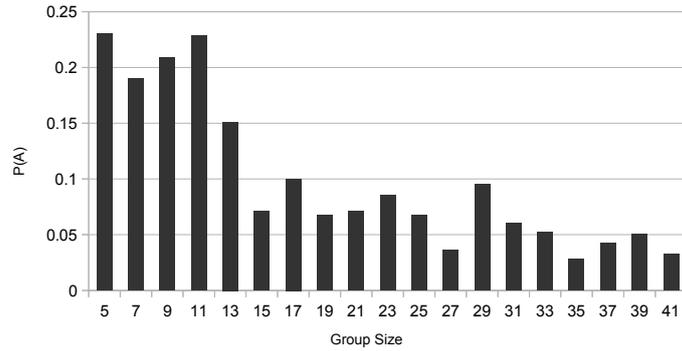


Figure 6: The probability of A as a function of the group size. The first speaker has a reliability of 0.5; the other group members have reliabilities randomly assigned from a uniform distribution.

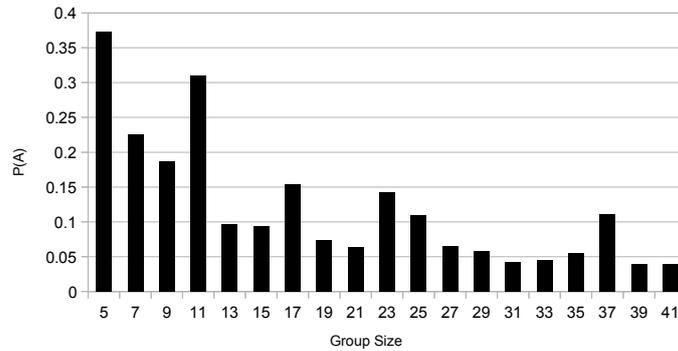
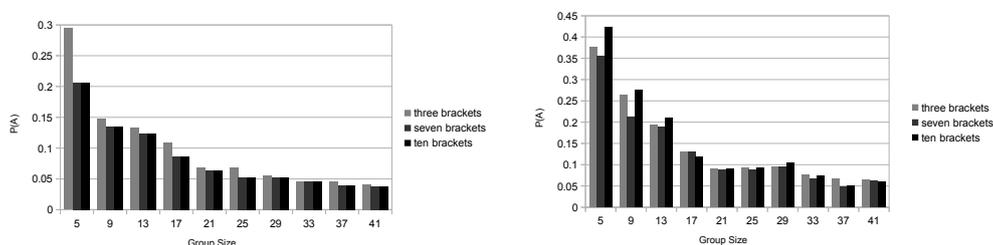


Figure 7: The probability of A as a function of the group size. The first speaker has a reliability of 0.5; the other group members have reliabilities randomly assigned from a β -distribution.

So far our simulations assumed that there are only three reliability brackets: H , M and L . One may wonder whether the anchoring effect is simply a consequence of this assumption which models the group members as not fully rational. Fully rational agents will take all available information (e.g. their assessment of the reliability of the other group members) into account. To investigate the consequences of this assumption, we will next look into the anchoring effect for groups with a higher number of reliability brackets.

The effect of increasing the number of reliability brackets depends on how the reliability of the first speaker compares to that of the majority of the group members. In general, and as one would expect, for groups where the first speaker is more reliable than the majority of the group members, increasing the number of reliability brackets increases

the anchoring effect while for groups with a less reliable first speaker increasing the number of reliability brackets has the opposite effect of reducing the anchoring effect. However, our simulations suggest that for group where the majority of group members have reliabilities close to that of the first speaker, increasing the number of reliability brackets does not make a significant effect, especially for larger groups. This is not all that surprising: If the other group members have reliabilities close to the reliability of the first speaker, increasing the number of reliability brackets will still keep them in a reliability bracket that is the same or close to that of the first speaker. Thus increasing the number of reliability brackets does not make a significant change in the weights that the other group members assign to the first speaker. Figures 8(a) and 8(b) show the dependence of the anchoring effect on the number of reliability brackets for two such groups. The plots suggest that although changing the number of reliability brackets can influence the anchoring effect in these groups in both directions, the changes are small and negligible for medium and large groups.

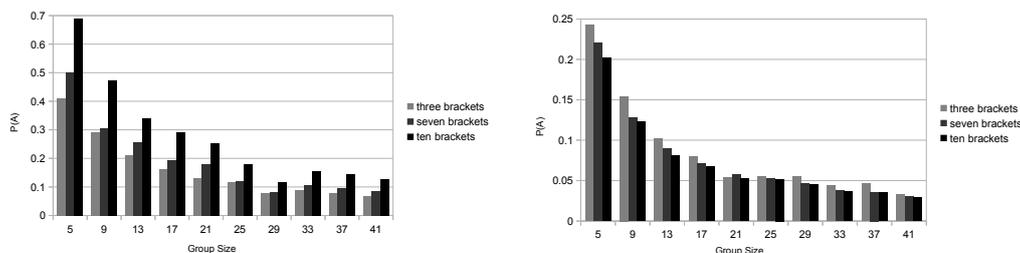


(a) The first speaker has a reliability of 0.55; (b) The first speaker has a reliability of 0.7; the other group members have reliabilities of 0.7 and 0.65 (equally distributed). the other group members have reliabilities of 0.5 and 0.65 (equally distributed).

Figure 8: The probability of A as a function of the group size for different numbers of reliability brackets.

In contrast, when there is a significant difference between the reliability of the first speaker and the reliability of the majority of the group members, increasing the number of reliability brackets makes a visible change to the anchoring effect. This will lead to a considerable increase of the anchoring effect when the first speaker is far more reliable than the majority and to its decrease when she is significantly less reliable. See for example Figure 9(a) where the first speaker has a reliability of 0.85 and the rest of the group have reliabilities of 0.4 and 0.6. Notice that for this group, when we restrict ourselves to three reliability brackets, the first speaker will, on average, be estimated as highly reliable (H) and most of the other group members will, on average, be perceived as having a medium reliability (M) value; the opinions will be weighted accordingly. When we move to a higher number of reliability brackets, the first speaker will still be estimated as highly reliable while at least half of the other group members will now be assigned a reliability value lower than medium; hence the first speaker will receive a higher relative weight. This will naturally increase the anchoring effect, as the plot shows. Figure 9(b) shows the reverse situation where the first speaker has a reliability of 0.4 while the rest

of the group have reliabilities of 0.6 and 0.85. As expected, increasing the number of reliability brackets will clearly reduce the anchoring effect. Notice, however, that in all these cases the anchoring effect persists; it is only the extent of the effect that changes by increasing the number of reliability brackets. This supports the claim that the anchoring effect is indeed a result of the order in which the group members express their opinions but the extent of the anchoring effect can be influenced by the the number of reliability brackets.



(a) The first speaker has a reliability of 0.85; the other group members have reliabilities of 0.4 and 0.6 (equally distributed). (b) The first speaker has a reliability of 0.4; the other group members have reliabilities of 0.6 and 0.85 (equally distributed).

Figure 9: The probability of A as a function of group size for different number of reliability brackets.

Finally, we study the effect of the second order reliability on the anchoring effect.⁸ A group member's second order reliability measures how good she is in assessing the (first order) reliabilities of the other group members. We first note that changing the second order reliability in an almost homogeneous group does not have much of an effect as any effect will average out over several rounds and several runs of the simulation. However, if the first order reliability of the first speaker differs significantly from the first order reliabilities of the other group members, then a change of the second order reliability will have a considerable effect similar to changing the number of reliability brackets as discussed above: If the first speaker is significantly more reliable than the majority of the group members, increasing the second order reliability will only make this difference more obvious to the other group members and hence result in the assignment of a higher weight to the first speaker by the other group members. This, in turn, makes the anchoring effect more pronounced. In contrast, if the first speaker is significantly less reliable than the majority of the other group members, increasing the second order reliability will allow the other group members to better assess the reliability of the first speaker. This will result in a lower weight assigned to the first speaker by the other group members, which in turn reduces the anchoring effect. Figure 10 illustrates this case. Here we consider a group of size 11 where the first speaker is highly reliable (reliability = 0.9), while the rest of the group has a rather low reliability of 0.3. We see that the anchoring effect increases with increasing second order reliability.

⁸Note that we assume throughout this study that all group members have the same second order reliability.

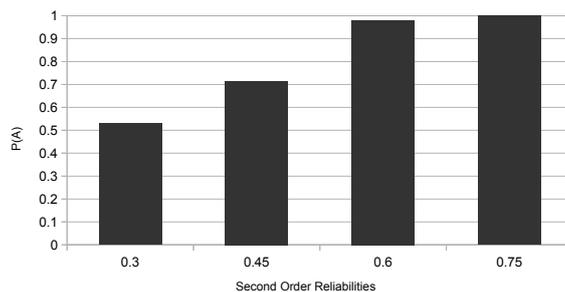


Figure 10: The probability of A as a function of the second order reliability for a group of size 11. The first speaker has a (first order) reliability of 0.85; the other group members have (first order) reliabilities of 0.3.

5 Conclusions

The anchoring effect shows up in many different contexts and has been studied in-depth in a wide range of experimental investigations. However, these works do not explain the occurrence of the effect. They do not answer the questions why and under which conditions the effect occurs. To make some progress towards addressing these questions, formal models can be of much help. They allow us to investigate under which conditions the effect occurs, and they allow us to study counterfactual scenarios that show what would happen under those conditions. This article makes an attempt in this direction. However, we do not claim to have provided a full explanation of the anchoring effect. Rather, our aim is much more modest: We want to understand *one aspect of one instance* of the anchoring effect by addressing the following question: does the anchoring effect in deliberations also occur in groups of individually rational agents? To do so, we constructed a model of the deliberation process, inspired by the Lehrer-Wagner model, and made a couple of assumptions about what it means for a group member to be individually rational in a deliberation process. Based on analytical calculations and computer simulations, we have shown that the answer to our question is yes, suggesting that the anchoring effect is even more robust than one thought. We have also seen that the effect does not disappear if one relaxes the (psychologically more realistic) assumption of considering only three reliability brackets. We take these findings to be progress towards a better understanding of the anchoring effect.

In future work, we plan to extend our studies in several directions. Here we mention only two of them:

1. Alternative models: Our results might still be an artefact of the idealizations made in the proposed model. We do not think so, as the effect occurs for a wide range of parameters and is stable (though less pronounced) if the number of reliability brackets is increased. And yet, it will be important to study variants of the present model (imposing, e.g., also network structures) and to explore the

possible occurrence of the anchoring effect in alternative modeling frameworks such as the suitably adapted Hegselmann-Krause model.

2. Interventions: It would be interesting to explore which constraints on (or designs of) the deliberation process help preventing the anchoring effect. This question can first be addressed in computer simulations, building on and extending the present model. Subsequently, experiments could be carried out on the basis of the simulation results to test them empirically.

Acknowledgment

We are grateful to Klaus Fiedler, Ulrike Hahn, Martin Kocher, Jan Sprenger and Anja Tuschke for helpful discussions and to the Alexander von Humboldt Foundation for financial support. S.H. would like to thank LMU Munich's Center for Advanced Studies (CAS) for support through the Senior Researcher in Residence program in the academic year 20016/17.

References

- Bazerman, M. (2002). *Judgement in Managerial Decision Making*. Chichester, UK: Wiley.
- Elga, A. (2007). Reflection and Disagreement. *Noûs* 41(3): 478–502.
- Frigg, R. & S. Hartmann (2006). Models in Science. *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition, revised for the Fall 2012 Edition).
- Hartmann, S., C. Martini & J. Sprenger (2009). Consensual Decision-Making Among Epistemic Peers. *Episteme* 6: 110–129.
- Hartmann, S. & S. Rafiee Rad (2017). Voting, Deliberation, and Truth. To appear in *Synthese*. DOI: 10.1007/s11229-016-1268-9.
- Hegselmann, R. & U. Krause (2002). Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation* 5(3).
- Hegselmann, R. & U. Krause (2006). Truth and Cognitive Division of Labour: First Steps Towards a Computer Aided Social Epistemology. *Journal of Artificial Societies and Social Simulation* 9(3).
- Hegselmann, R. & U. Krause (2009). Deliberative Exchange, Truth, and Cognitive Division of Labour: A Low-Resolution Modeling Approach. *Episteme* 6: 130–144.
- Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, & A. Stone (2006). Would you be Happier if you Were Richer? A Focusing Illusion. *Science* 312(5782): 1908–1910.

- Lehrer, K. (1976). When Rational Disagreement is Impossible. *Nou̇s* 10: 327–332.
- Lehrer, K. & C. Wagner (1981). *Rational Consensus in Science and Society*. Dordrecht: Reidel.
- Mussweiler, T., B. Englich & F. Strack (2004). Anchoring Effect, in: R. Pohl (ed.): *Cognitive Illusions: A Handbook of Fallacies and Biases in Thinking, Judgement, and Memory*. Psychology Press, pp. 183–196.
- Olsson, E.J. (2011). A Bayesian Simulation Model of Group Deliberation and Polarization. In: F. Zenker (ed.): *Bayesian Argumentation*. Berlin: Springer.
- Oreskes, N. (2004). The Scientific Consensus on Climate Change. *Science* 306(4): 1986.
- Reutlinger, A. (2016). Is There A Monist Theory of Causal and Non-Causal Explanations? The Counterfactual Theory of Scientific Explanation. *Philosophy of Science* 83: 733–745.
- Reutlinger, A., D. Hangleiter & S. Hartmann. Understanding (with) Toy Models. To appear in *The British Journal for the Philosophy of Science*.
- Steele, K., H. Regan, M. Colyvan & M. Burgman. Right Decisions or Happy Decision-Makers? *Social Epistemology: A Journal of Knowledge, Culture and Policy* 21: 349–368.
- Tversky, A. & D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* 185(4157): 1124–1131.

Appendix A: Proof of Theorem 2

We first prove the following lemma.

Lemma 1 For $1 \leq m \leq n$,

$$B_n^m \cdot B_n^{m-1} \cdot \dots \cdot B_n^1 = \begin{pmatrix} \bar{B}^m & \mathbf{0} \\ C^m & I_{n-m} \end{pmatrix},$$

where

$$\bar{B}^m := \begin{pmatrix} 2^{m-1} + 1 & 2^{m-2} & \dots & 2^0 \\ 2^{m-1} & 2^{m-2} + 1 & \dots & 2^0 \\ 2^{m-1} & 2^{m-2} & \dots & 2^0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 2^{m-1} & 2^{m-2} & \dots & 2^0 + 1 \end{pmatrix}_{m \times m}.$$

$\mathbf{0}$ is the $m \times (n - m)$ zero matrix, I_{n-m} is the $(n - m) \times (n - m)$ identity matrix and

$$C^m := \begin{pmatrix} 2^{m-1} & 2^{m-2} & \dots & 2^0 \\ 2^{m-1} & 2^{m-2} & \dots & 2^0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 2^{m-1} & 2^{m-2} & \dots & 2^0 \end{pmatrix}_{(n-m) \times m}.$$

Proof:

We proof the lemma by induction. For $m = 1$ the result is trivially true. Next, we suppose that it is true for $m \geq 1$ and show that it also holds for $m + 1$.

$$\begin{aligned} B_n^{m+1}.B_n^{m-1}.\dots.B_n^1 &= B_n^{m+1} \cdot \begin{pmatrix} \bar{B}^m & \mathbf{0} \\ C^m & I_{n-m} \end{pmatrix} \\ &= (A_n^{m+1} + I_n) \cdot \begin{pmatrix} \bar{B}^m & \mathbf{0} \\ C^m & I_{n-m} \end{pmatrix} \\ &= A_n^{m+1} \cdot \begin{pmatrix} \bar{B}^m & \mathbf{0} \\ C^m & I_{n-m} \end{pmatrix} + I_n \cdot \begin{pmatrix} \bar{B}^m & \mathbf{0} \\ C^m & I_{n-m} \end{pmatrix} \\ &= \begin{pmatrix} \bar{B}^{m+1} & \mathbf{0} \\ C^{m+1} & I_{n-(m+1)} \end{pmatrix}. \end{aligned}$$

This completes the proof of Lemma 1. ■

Corollary 1 *The result of updating the assignments $\vec{V}^{(0)}$ through one round of deliberation is given by*

$$\begin{aligned} \vec{V}^{(1)} &= \frac{1}{2^n} (B_n^n.B_n^{n-1}.\dots.B_n^1) \vec{V}^{(0)} \\ &= \left(\begin{pmatrix} 2^{n-1} & 2^{n-2} & \dots & 2^0 \\ 2^{n-1} & 2^{n-2} & \dots & 2^0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 2^{n-1} & 2^{n-2} & \dots & 2^0 \end{pmatrix} + \frac{1}{2^n} I_n \right) \vec{V}^{(0)}. \end{aligned}$$

Proposition 1 *Let*

$$B := \frac{1}{2^n} \prod_{i=1}^n B_n^i - \frac{1}{2^n} I_n,$$

then the result of updating the assignments $\vec{V}^{(0)}$ through k rounds of deliberation is given by

$$\begin{aligned}\vec{V}^{(k)} &= \left(\frac{1}{2^n} (B_n^n \cdot B_n^{n-1} \dots B_n^1) \right)^k \vec{V}^{(0)} \\ &= \left(\sum_{t=1}^k \binom{k}{t} \frac{(2^n - 1)^{t-1}}{(2^n)^{k-1}} B + \frac{1}{2^{nk}} I_n \right) \cdot \vec{V}^{(0)}.\end{aligned}$$

Proof:

Let $b_i := \sum_{j=1}^n \langle B \rangle_{ij}$. Notice that by Lemma 1, in matrix B all rows are equal so $b_1 = b_2 = \dots = b_n = b$, where $b := \sum_{i=1}^n 2^{-i} = 1 - \frac{1}{2^n}$. Moreover, we have $B^k = b^{k-1} B$. Hence,

$$\begin{aligned}\vec{V}^{(k)} &= \left(\frac{1}{2^n} \prod_{i=1}^n B_n^i \right)^k \vec{V}^{(0)} = \left(B + \frac{1}{2^n} I_n \right)^k \vec{V}^{(0)} \\ &= \left(\frac{1}{2^{nk}} I_n + \sum_{t=1}^k \binom{k}{t} \left(\frac{1}{2^n} I_n \right)^{k-t} B^t \right) \vec{V}^{(0)} \\ &= \left(\frac{1}{2^{nk}} I_n + \sum_{t=1}^k \binom{k}{t} \left(\frac{1}{2^n} I_n \right)^{k-t} (b^{t-1} B) \right) \vec{V}^{(0)} \\ &= \left(\frac{1}{2^{nk}} I_n + \sum_{t=1}^k \binom{k}{t} \frac{(1 - \frac{1}{2^n})^{t-1}}{2^{n(k-t)}} B \right) \vec{V}^{(0)} \\ &= \left(\frac{1}{2^{nk}} I_n + \sum_{t=1}^k \binom{k}{t} \frac{(2^n - 1)^{t-1}}{2^{n(k-1)}} B \right) \vec{V}^{(0)},\end{aligned}$$

which completes the proof. ■

Proof of Theorem 2

Let \vec{V} be the asymptotic result of the deliberation process and let w_i be the weight assigned to the i^{th} speaker in the limit. By Proposition 1,

$$\begin{aligned}\vec{V} &= \lim_{k \rightarrow \infty} \left(\sum_{t=1}^k \binom{k}{t} \frac{(2^n - 1)^{t-1}}{(2^n)^{k-1}} B + \frac{1}{2^{nk}} I_n \right) \vec{V}^{(0)} \\ &= \lim_{k \rightarrow \infty} \left(\sum_{t=1}^k \binom{k}{t} \frac{(2^n - 1)^{t-1}}{(2^n)^{k-1}} \right) B \vec{V}^{(0)} \\ &= d B V^{(0)},\end{aligned}\tag{8}$$

where $d := \lim_{k \rightarrow \infty} \sum_{t=1}^k \binom{k}{t} \frac{(2^n - 1)^{t-1}}{(2^n)^{k-1}}$. Then, by Proposition 1, the weight assigned to the i^{th} speaker is given by $w_i = d \cdot 2^{n-i}$. Hence, $w_1 \geq w_2 \geq \dots \geq w_n$. ■

Appendix B: Stability Result

In this appendix we will investigate the stability of our results for the homogeneous groups by showing that the results still hold for small changes in the equal-weight assumption. Here the group members consider small deviation from the equal weights by assigning $\frac{1-\epsilon}{2}$ to the speaker and $\frac{1+\epsilon}{2}$ to themselves. As before, after the i^{th} group member speaks, everyone will update their assignment as a weighted average of their current value and that announced by the i^{th} speaker while assigning a slightly higher/lower weight to themselves as opposed to the speaker. This process can again be represented by matrix multiplication as before by replacing the matrices B_n^i with

$$B_n^{i'} = B_n^i - \epsilon E_n^i$$

where

$$E_n^j = B_n^j - 2A_n^j.$$

Thus, for example, in a group of size 3, the matrix

$$B_3^2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

will be replaced by

$$\begin{aligned} B_3^{2'} &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} + \epsilon \left(\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} - 2 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right) \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} + \epsilon \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1+\epsilon & 1-\epsilon & 0 \\ 0 & 2 & 0 \\ 0 & 1-\epsilon & 1+\epsilon \end{pmatrix}. \end{aligned}$$

With this, we obtain the following result after one round of deliberation:

$$\vec{V}^{(1)} = \frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \cdot \vec{V}^{(0)}$$

Similarly, after k rounds of deliberation we shall have

$$\vec{V}^{(k)} = \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)}.$$

We are now in the position to prove the corresponding versions of Lemma 1 for this setting.

Lemma 2 Let A_n^i 's be the matrices with 1 on the i^{th} column and zero elsewhere and $B_n^j = I_n + A_n^j$ as before. Then

$$\prod_{k=j+1}^n B_n^k A_n^j \prod_{k=1}^{j-1} B_n^k = 2^{n-j} A_n^j + \sum_{k=1}^{j-1} 2^{n-1-k} A_n^k.$$

Proof.

$$\prod_{k=j+1}^n B_n^k A_n^j \prod_{k=1}^{j-1} B_n^k = \prod_{k=j+1}^n (I_n + A_n^k) A_n^j \prod_{k=1}^{j-1} B_n^k.$$

First notice that for all $i, j = 1, \dots, n$ we have $A_n^i \cdot A_n^j = A_n^j$, thus

$$\prod_{k=j+1}^n B_n^k A_n^j \prod_{k=1}^{j-1} B_n^k = \prod_{k=j+1}^n (I_n + A_n^k) A_n^j \prod_{k=1}^{j-1} B_n^k = 2^{n-j} A_n^j \prod_{k=1}^{j-1} B_n^k.$$

From Lemma 1 we have

$$\prod_{k=1}^{j-1} B_n^k = I_n + \sum_{k=1}^{j-1} 2^{j-1-k} A_n^k,$$

thus

$$\begin{aligned} \prod_{k=j+1}^n B_n^k A_n^j \prod_{k=1}^{j-1} B_n^k &= \prod_{k=j+1}^n (I_n + A_n^k) A_n^j \prod_{k=1}^{j-1} B_n^k \\ &= 2^{n-j} A_n^j \left(I_n + \sum_{k=1}^{j-1} 2^{j-1-k} A_n^k \right) \\ &= 2^{n-j} A_n^j + 2^{n-j} A_n^j \sum_{k=1}^{j-1} 2^{j-1-k} A_n^k \\ &= 2^{n-j} A_n^j + \sum_{k=1}^{j-1} 2^{n-1-k} A_n^k \end{aligned}$$

as required. ■

Next we have a modified version of the Theorem 1:

Proposition 2

$$\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} = \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \epsilon \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) + \frac{O(\epsilon^2)}{2^n}$$

Proof: $\frac{1}{2^n} \prod_{i=1}^n B_n^i =$

$$\begin{aligned}
&= \frac{1}{2^n} \prod_{i=1}^n (B_n^i + \epsilon E_n^i) \\
&= \frac{1}{2^n} \left(\prod_{i=1}^n B_n^i + \epsilon \sum_{j=1}^n \left(\prod_{k=j+1}^n B_n^k \right) E_n^j \left(\prod_{k=1}^{j-1} B_n^k \right) + O(\epsilon^2) \right) \\
&= \frac{1}{2^n} \left(\prod_{i=1}^n B_n^i + \epsilon \sum_{j=1}^n \left(\prod_{k=j+1}^n B_n^k \right) (B_n^j - 2A_n^j) \left(\prod_{k=1}^{j-1} B_n^k \right) + O(\epsilon^2) \right) \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(\prod_{k=1}^n B_n^k - 2 \prod_{k=j+1}^n B_n^k A_n^j \prod_{k=1}^{j-1} B_n^k \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(\prod_{k=1}^n B_n^k - 2 \left(2^{n-j} A_n^j + \sum_{k=1}^{j-1} 2^{n-1-k} A_n^k \right) \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(\left(I_n + \sum_{k=1}^n 2^{n-k} A_n^k \right) - \left(2^{n-j+1} A_n^j + \sum_{k=1}^{j-1} 2^{n-k} A_n^k \right) \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(I_n + \sum_{k=j}^n 2^{n-k} A_n^k - 2^{n-j+1} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(\sum_{k=j}^n 2^{n-k} A_n^k - 2^{n-j+1} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \frac{\epsilon}{2^n} \sum_{j=1}^n \left(\sum_{k=j+1}^n 2^{n-k} A_n^k - 2^{n-j} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \frac{\epsilon}{2^n} \left(\sum_{j=1}^n \left(\sum_{k=j+1}^n 2^{n-k} A_n^k \right) - \left(\sum_{j=1}^n 2^{n-j} A_n^j \right) \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \frac{\epsilon}{2^n} \left(\sum_{j=1}^n (j-1) 2^{n-j} A_n^j - \sum_{j=1}^n 2^{n-j} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \frac{\epsilon}{2^n} \left(\sum_{j=1}^n (j-2) 2^{n-j} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \frac{n\epsilon}{2^n} I_n + \epsilon \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) + \frac{O(\epsilon^2)}{2^n} \\
&= \frac{1}{2^n} \prod_{i=1}^n B_n^i + \epsilon \begin{pmatrix} \frac{-1}{2} + \frac{n}{2^n} & 0 & \frac{1}{2^3} & \frac{2}{2^4} & \cdots & \frac{n-2}{2^n} \\ \frac{-1}{2} & 0 + \frac{n}{2^n} & \frac{1}{2^3} & \frac{2}{2^4} & \cdots & \frac{n-2}{2^n} \\ \frac{-1}{2} & 0 & \frac{1}{2^3} + \frac{n}{2^n} & \frac{2}{2^4} & \cdots & \frac{n-2}{2^n} \\ \cdot & \cdot & 25 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{-1}{2} & 0 & \frac{1}{2^3} & \frac{2}{2^4} & \cdots & \frac{n-2}{2^n} + \frac{n}{2^n} \end{pmatrix} + \frac{O(\epsilon^2)}{2^n}
\end{aligned}$$

where the fifth equality is given by Lemma 2. ■

We are now in the position to prove the equivalent of Theorem 2 for the agents that can make small deviation from the equal weight assumptions.

Theorem 4 *For ϵ sufficiently small, the process of deliberation described above with weights set as $\frac{1-\epsilon}{2}$ and $\frac{1+\epsilon}{2}$ will end in consensus. That is if*

$$\vec{V} = \lim_{k \rightarrow \infty} \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)}$$

then $\langle \vec{V} \rangle_i = \langle \vec{V} \rangle_j$ for $i, j = 1, \dots, n$. Moreover in the final consensus the opinion of the first speaker receives the highest weight compared to other group members.

Proof.

Notice that $\frac{1}{2^n} \prod_{i=1}^n B_n^{i'}$ is essentially a weight matrix in the sense of Lehrer-Wagner model and thus the convergence in the limit follows for the same reason as in the Lehrer-Wagner model. We will now show that the first speaker receives the highest weight in the final consensus. Let $C_n := \frac{1}{2^n} \prod_{i=1}^n B_n^i$ and $\mathcal{O}_n^\epsilon := \frac{O(\epsilon^2)}{2^n}$, by Proposition 2 we have

$$\begin{aligned} \vec{V} &= \lim_{k \rightarrow \infty} \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)} \\ &= \lim_{k \rightarrow \infty} \left(C_n + \frac{n\epsilon}{2^n} I_n + \epsilon \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) + \mathcal{O}_n^\epsilon \right)^k V^{(0)} \\ &= \lim_{k \rightarrow \infty} \left(\sum_{n_1+n_2+n_3+n_4=k} C_n^{n_1} \left(\frac{n\epsilon}{2^n} I_n \right)^{n_2} \left(\epsilon \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right)^{n_3} (\mathcal{O}_n^\epsilon)^{n_4} \right) V^{(0)}. \end{aligned} \tag{9}$$

Any term with $n_2 \geq 2$ or $n_3 \geq 2$ or $n_4 \geq 1$ include a term in the order of $O(\epsilon^2)$ and thus can be ignored, the same is true for the term with $n_2 = n_3 = 1$ so

$$\begin{aligned}
\vec{V} &= \lim_{k \rightarrow \infty} \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)} \\
&= \lim_{k \rightarrow \infty} \left(C_n^k + C_n^{k-1} \left(\frac{n\epsilon}{2^n} I_n \right) + C_n^{k-1} \left(\epsilon \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \right) \vec{V}^{(0)} \\
&= \lim_{k \rightarrow \infty} C_n^k \vec{V}^{(0)} + \lim_{k \rightarrow \infty} C_n^{k-1} \left(\frac{n\epsilon}{2^n} I_n + \epsilon \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \vec{V}^{(0)} \\
&= dB \vec{V}^{(0)} + dB \left(\frac{n\epsilon}{2^n} I_n + \epsilon \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \vec{V}^{(0)} \\
&= dB \left(\left(1 + \frac{n\epsilon}{2^n} \right) I_n + \epsilon \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \vec{V}^{(0)} \\
&= d \left(1 + \frac{n\epsilon}{2^n} \right) B \vec{V}^{(0)} + d\epsilon (B \cdot \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j) \vec{V}^{(0)}
\end{aligned}$$

where $d = \lim_{k \rightarrow \infty} \sum_{t=1}^k \binom{k}{t} \frac{(2^n-1)^{t-1}}{(2^n)^{k-1}}$ and $B = \frac{1}{2^n} \prod_{i=1}^n B_n^i - \frac{1}{2^n} I_n$ as in Theorem 2. Notice that B and $\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j$ are both matrices with all rows equal thus

$$B \cdot \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j = \left(\sum_{k=1}^n B_{ik} \right) \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) = \left(1 - \frac{1}{2^n} \right) \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right)$$

where the last equality is derived from Corollary 1. Thus

$$\begin{aligned}
\vec{V} &= \lim_{k \rightarrow \infty} \left(\frac{1}{2^n} \prod_{i=1}^n B_n^{i'} \right)^k \vec{V}^{(0)} \\
&= c \left(d \left(1 + \frac{n\epsilon}{2^n} \right) B + d\epsilon \left(B \cdot \sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \right) \vec{V}^{(0)} \\
&= d \left(\left(1 + \frac{n\epsilon}{2^n} \right) B + \epsilon \left(1 - \frac{1}{2^n} \right) \left(\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j \right) \right) \vec{V}^{(0)}
\end{aligned}$$

The weight assigned to the i^{th} member will be the sum of the weights assigned through B and $\sum_{j=1}^n \frac{(j-2)}{2^j} A_n^j$ that is

$$d \left(1 + \frac{n\epsilon}{2^n} \right) 2^{-i} + d\epsilon \left(1 - \frac{1}{2^n} \right) \frac{(i-2)}{2^i}$$

For n large enough we have $\frac{n\epsilon}{2^n} \approx 0$ and hence w_i (the weight assigned to the i^{th} speaker) will be

$$d \cdot 2^{-i} + d\epsilon \left(1 - \frac{1}{2^n}\right) \frac{(i-2)}{2^i}. \quad \blacksquare$$

It is now easy to check that for $\epsilon \leq 1/2$ we have $w_1 \geq w_i$ for $i = 1, \dots, n$; that means the first speaker receives the highest weight.

Appendix C: The Algorithm for the Simulation Procedure

The general sketch of the algorithm used for simulation is as follows:

1. First we set the relevant parameters. This includes the number of group members, n , the first order reliabilities, r_i , initial second order reliabilities, c_i as well as the maximum second order reliabilities C_i . We set the number of simulation runs to $N = 10^5$ and limit the number of deliberation rounds to $K = 40$. Unless otherwise specified, we use the assignment: $c_i = 0.8$ and $C_i = 0.9$ for all $i = 1, \dots, n$.
2. The simulation proceeds as follows.
 - (a) Initialize the prior values assigned by the group members randomly. (b) Compute the estimated reliabilities r_{ij} from a uniform distribution around the value r_j using eq. (4). (c) Calculate the estimated reliability brackets. (d) Start deliberation steps $1, \dots, n$: in step i update all assignments and reliability rankings based on the assignment of the i^{th} member and her reliability. (e) If a consensus is reached, stop. If there is no consensus yet, repeat the deliberation steps $1, \dots, n$. (f) Repeat (e) at most K times. (g) If a consensus is reached, calculate the difference between the consensus value and the initial assignment for each group member. (h) Add 1 to a counter if the consensus is closest to the initial value assigned by the first group member. Do nothing otherwise. (i) After N runs, compute the probability that the consensus is closest to the initial assignment of the first group member.