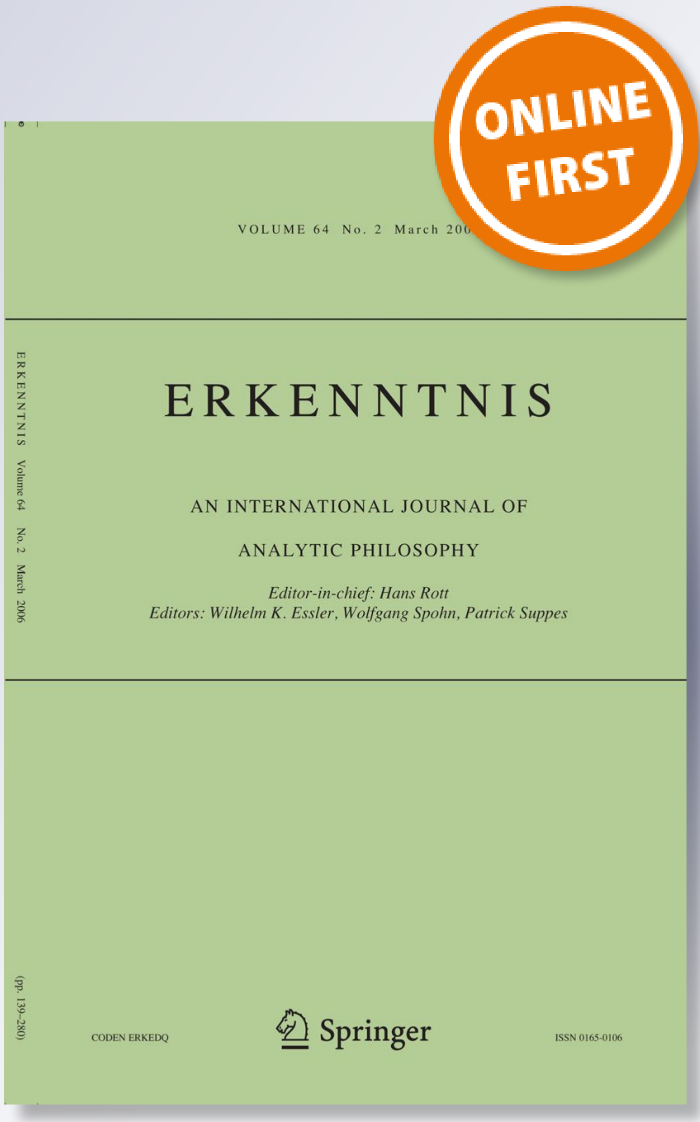
Resolving Disagreement Through Mutual Respect

Carlo Martini, Jan Sprenger & Mark Colyvan

Erkenntnis
An International Journal of Scientific Philosophy

ISSN 0165-0106

Erkenn
DOI 10.1007/s10670-012-9381-8



Erkenn
DOI 10.1007/s10670-012-9381-8

Resolving Disagreement Through Mutual Respect

Carlo Martini · Jan Sprenger · Mark Colyvan

Received: 9 August 2011 / Accepted: 11 May 2012

Abstract This paper explores the scope and limits of rational consensus through mutual respect, with the primary focus on the best known formal model of consensus: the Lehrer–Wagner model. We consider various arguments against the rationality of the Lehrer–Wagner model as a model of consensus about factual matters. We conclude that models such as this face problems in achieving rational consensus on disagreements about unknown factual matters, but that they hold considerable promise as models of how to rationally resolve non-factual disagreements.

1 Introduction: Consensus Versus Compromise

The topic of resolving disagreement has recently been receiving a great deal of attention, particularly in epistemology. From a pluralist perspective, disagreement can be seen as epistemically and socially desirable. The diversity of opinion characteristic of disagreement can, for example, prompt alternative lines of research in science, which, in turn, may be more fruitful in the long run than existing research programs. Having such a pluralist attitude seems epistemically sound since, in most practical cases, there is no a-priori method of assessing whether a particular line of

C. Martini (✉)
Philosophy and Economics Program, Institut für Philosophie, University of Bayreuth,
95440 Bayreuth, Germany
e-mail: carlo.martini@uni-bayreuth.de

J. Sprenger
Tilburg Center for Logic and Philosophy of Science, Tilburg University, Tilburg,
The Netherlands
e-mail: j.sprenger@uvt.nl

M. Colyvan
Sydney Centre for the Foundations of Science, University of Sydney, Sydney, NSW 2006, Australia
e-mail: mark.colyvan@sydney.edu.au

research will pay off: Armstrong (2001) and Page (2007), among others, argue along those lines, suggesting that groups perform better epistemically, when they take into account diverse sources of information and employ a variety of methods.

There are also social reasons to embrace disagreement. It might be argued that having dissident voices in a society is important for the long-term health of the society—minority political parties can help fuel public debate and keep the majority parties on their toes. Moreover, disagreement can be seen as an antidote to *group think*, and licenses a spreading of epistemic and social risks. Let a hundred flowers bloom, both epistemically and socially.

There is a down side to disagreement though. Disagreement can delay or even block group decision-making processes, where a group has to concur on a common course of action. If groups need to choose a single course of action, they have to settle their disagreements in one way or the other. This situation is normally recognized also by those who defend the epistemic gains from diversity. For instance, Hong and Page recognize that "diverse groups often have more conflict, more problems with communication, and less mutual respect and trust among members" (Hong and Page 2004, 16386). In other words, diversity has its downsides, and some times convergence is needed, either by voting, by compromising, or by a consensual resolution.

In a compromise situation, group members consent to a procedure of resolving the disagreement, but do not necessarily endorse the outcome delivered by this procedure. The agents in the group may accept the final outcome as the group decision, but privately hold firm to their original divergent views. Contrast this with a true consensus, where the agents adopt the final outcome as their own position on the matter in question. Such a consensual resolution has many attractive features.

When consensus is achieved, all agents have genuinely changed their views to the consensual view and the group opinion coincides with the opinion of every individual. Everybody gets their way when there is consensus. On the other hand, with compromise no individual changes their opinion. Here the group opinion need not be the opinion of anyone in the group and, in a sense, everyone can be dissatisfied with the group outcome. Apart from the fact that individuals *prefer* to agree with the rest of the group (Steele et al. 2007), a consensus reinforces the cohesion of the group. And of course achieving a consensus enables the group to function as a single agent. Finally, consensual resolutions are likely to be more *stable* than compromises. Decisions arrived at by compromise might be challenged on the next occasion by those dissatisfied with the result. Such considerations justify the efforts put into deliberation procedures that aim at finding a common ground on which all group members agree. It is clear, then, that in at least some circumstances, there are reasons to pursue consensus. So it is natural to ask how the emergence of consensus can be described and modeled.

In the following sections, we discuss the emergence of consensus through *mutual respect* among the group members. That is, we consider whether models of consensus through respect-based opinion convergence can be thought of as the result of a rational consensus process, submitted to by some suitably idealized agents. We start with a discussion of what it means to be a rational model of consensus and subsequently conduct a case study on the Lehrer–Wagner model

(Lehrer 1976), a salient representative of respect-based consensus models. The Lehrer–Wagner model has been criticized on several fronts as a model about factual disagreement. We reformulate and extend those criticisms, and argue that the model has descriptive and normative shortcomings when applied to *epistemic* disagreements. However, we argue that the model fares much better when used to resolve non-factual disagreements, such as disagreements on preferences, ethical or aesthetic values. We conclude that convergence of opinions through mutual respect is a valid strategy for non-factual disagreements. Such a conclusion is an important one, insofar as rational consensus in, for example, group agency and public policy matters involves matters of value, preferences, and other non-epistemic factors.

2 Models of Rational Consensus

There are two major questions about reaching a consensus. First, why should we be in a state of consensus rather than in a state of disagreement (Christensen 2007; Elga 2007; Kelly 2005)? Second, what makes a consensus procedure a model of *rational* consensus, as opposed to being merely a pooling algorithm?

We have already sketched a response to the first question. Now, we focus on the second question. Whether a group adopts a final judgment as a consensus, or as a mere compromise, depends on whether individuals in the group have changed their attitudes and whether such changes can be rationally motivated. In particular, consensus models are distinguished from compromise models by their ability to provide a plausible story about why the process of arriving at an agreement might be understood as a—not necessarily unique—*rational attitude revision procedure*.

The rationality of a consensus model of attitude revision can be defended or criticized in many ways. Since most models contain a number of idealizations, we believe that the parts of a model only need to be rationally motivated in so far as they are crucial to the model as a whole. We will use this approach in our assessment of the Lehrer–Wagner model in Sects. 4 and 5. On the other hand, it is equally clear that we need more than a "non-irrationality condition", namely, a *positive philosophical account* of why we should accept a consensus model, and possibly only that model (this is the requirement of uniqueness, which will be treated in Sect. 4).

It is natural to think of consensus emergence and deliberation as a gradual process where *mutual respect* plays an important role. That is what happens in the Lehrer–Wagner model: agents reach consensus on a particular question from a set of diverging opinions by repeating one and the same respect-based procedure until the group members finally agree. The model was first developed as a descriptive mathematical model of group power relations in French (1956) and as a general model of consensus formation in DeGroot (1974). While DeGroot intended his model to be normative, his research focused on the fundamental mathematical properties of the model, leaving the interpretation, further elaboration, and philosophical justification to Lehrer and Wagner (1981).

The model tackles the problem of estimating a particular quantity x, from the individual estimates v_i of every group member i. This quantity x is normally thought of as objective and independent of the group members' cognitive states. The quantity x in dispute might, for instance, be the number of product sales estimated by a manufacturing firm, or the number of rainy days in a certain time frame.

Lehrer and Wagner's central idea consists in ascribing the agents beliefs about each other's expertise, or in other words, mutual assignments of respect as epistemic agents on the issue at hand. These weights of respect, w_{ij}, are the weight of respect agent i has for agent j and describe the proportion to which j's opinion on the subject matter in question affects i's revised opinion. These mutual respect assignments are used to revise the original estimates of the quantity in question, and codified in an $N \times N$ matrix W (where N denotes the number of agents in the group):

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{pmatrix}.$$

An important mathematical constraint is that the values in each row are nonnegative and normalized so as to sum to 1: $\sum_{j=1}^{N} w_{ij} = 1$. Thus, the w_{ij} represent relative weights which the agents ascribe to themselves and to others when it comes to estimating the unknown value x.[1] Then, W is multiplied with a vector v that contains the agents' individual estimates of x, obtaining a novel updated value for v:

$$W \cdot v = \begin{pmatrix} w_{11}v_1 + w_{12}v_2 + \cdots + w_{1N}v_N \\ w_{21}v_1 + w_{22}v_2 + \cdots + w_{2N}v_N \\ \cdots \\ w_{N1}v_1 + w_{N2}v_2 + \cdots + w_{NN}v_N \end{pmatrix}.$$

In general, however, this procedure will not directly lead to consensus, since the entries of $W \cdot v$ differ: $(Wv)_i \neq (Wv)_j$. In the original version of the model, it is argued that the agents reach a rational consensus by repeating the procedure until the opinions do not change any more. One possible interpretation is that at each round, agents use their list of weights (interpreted as respect assignments) in order to update their opinions, starting from the opinion that the group, as a whole, has reached up to the previous stage of updates. This allows the group members to repeat the averaging procedure until the disagreement is finally resolved (Wagner 1978). Clearly, this mechanism makes the Lehrer–Wagner model a paradigmatic representative of gradual opinion convergence through mutual respect: it is a function of original estimates and respect assignments, and it is gradual in the sense that, in most cases, a single application of the aggregation procedure is not sufficient to ensure convergence.

Lehrer and Wagner (1981) show that under very weak constraints, the sequence $(W^k)_{k \in \mathbb{N}}$ converges to a matrix W^∞ where all rows are identical, that is, where all agents agree on their relative weights (see also Jackson and Golub 2007). In other words, the agents do not only achieve consensus on the issue under consideration (i.e., all entries of $W^\infty v$ are equal), but they also agree on the effective relative

[1] These weights can but need not depend on the closeness of the agents' estimates as in Regan et al. (2006) and Hegselmann and Krause (2002).

weight of each group member, that is, each other's expertise. This is, apparently, the strongest possible form of consensus, but it is less clear that such a strong form of consensus is necessary or even desirable. We need an account of what makes such a procedure rational, and we discuss this issue in the following three sections.

3 Factual Consensus Through Mutual Respect

In factual disagreements, models of consensus through mutual respect describe an agent's final judgment as a function of two pieces of information: (1) the mutual assessment of the degrees of expertise and (2) each member's individual (and presumably independent) judgment on the subject matter under deliberation. The Lehrer–Wagner perfectly conforms to this schema, as the previous section made clear. Now, we can ask the question of whether such models of factual disagreement resolution are rational. Lehrer (1976) provides an argument from consistency: once agents accept to enter into deliberation, they implicitly assign relative weights to others. If the agents do not give at least *some* weight to the other agents, their behavior would amount to pure dogmatism.

> Actual disagreement among experts must result either from an incomplete exchange of information, individual dogmatism, or a failure to grasp the mathematical implications of their initial stage. What is impossible is that the members of some community of inquiry should grasp the mathematical implications of their initial state and yet disagree (Lehrer 1976, 331).

And again:

> One justification for aggregation is consistency, since refusing to aggregate is equivalent to assigning everyone else a weight of zero and aggregating. (Lehrer and Wagner 1981, 43)

In sum, the best possible formulation of the standard argument in defense of the Lehrer–Wagner model is that its mathematical structure captures a *typical* situation of disagreement. Refusing to change one's opinion would be equivalent, in mathematical terms, to assigning a null weight to all other members and full weight to oneself. This situation, which could prevent the model from reaching a consensual value, is one of pure dogmatism. It amounts to refusing to enter into the consensus process.

This line of defense may not hold water though: only those agents who have already accepted the assumptions of the model are rationally required to assign non-zero weights to the others, and to participate in the aggregation process. For example, Thomas Kelly has argued for a form of "epistemic egoism without apology" (Kelly 2005, 192). It might be argued that a situation of disagreement between epistemic peers does not necessarily commit one to revise one's belief unless there is *independent* evidence that one is wrong and our peer is right. Furthermore, even if an agent decides to respect the opinion of her peers, she is not obliged to repeat the process. After some rounds of pooling, all relevant information seems to be taken into account, such that repeating the procedure would only introduce additional biases and reward stubbornness. Instead, we might agree to disagree, and give up our goal of reaching a consensus. In deadlocked disagreements, our social goals (consensus) and our epistemic goals (truth) are clearly incommensurable and not amenable to a universally rational resolution. Lehrer's argument for consensual resolutions of disagreement therefore amounts to a *petitio principii*, exception made for his *moral* stance against dogmatism. A moral stance, however, is not enough to justify a particular form of aggregation as *epistemically* rational.

There is, however, a prima facie plausibility to the idea that the Lehrer–Wagner model is a model of rational consensus. This interpretation of the model hangs on appreciating the epistemic tension an agent finds herself in when (a) she admits that some other agent should have a say on the issue of disagreement (she gives that agent a positive weight) (b) she disagrees with the other agent on the matter at issue. The situation is akin to asserting "*P* but my peer says ¬*P*, and my peer is usually right about such things".[2] Similarly, consider holding the belief that someone is a very unreliable source of information, yet you find yourself agreeing with her. What should an agent do, when she finds herself disagreeing with someone whom she generally takes to be reliable, or conversely, when she agrees with someone she takes to be unreliable? Clearly, one would reasonably argue, she ought to revise either her opinion about the disputed matter or about the reliability of the other agent.

So, to return to the Lehrer–Wagner framework, it is (Moore-) paradoxical to assert that "my value for the disputed quantity is v_i, but someone whom I trust on such matters, and to whom I give a high weight of respect, disagrees with me". If we think of each group member as a partially reliable information channel, we have to reconcile the tensions between an agent's own view, the various channels, and the reliability assessments of those channels. The agent can either revise her weights of respect or her value of the disputed quantity. The Lehrer–Wagner model—and actually, all models where mutual respect triggers a convergence of opinions—take the second option, advising an agent to revise her judgments about the disputed quantity in light of her assessment of the reliability of the various information channels. If disagreement persists after each agent performs such a revision, the process should be repeated as many times as necessary. The convergence guaranteed by the model assures us that there is a stable, tension-free belief structure for this group.

It is surely rational to aim at relieving Moorean epistemic tensions in some such fashion. Especially if one realizes that the alternative—to relieve the tension by assigning zero weights to the other agents—may be epistemically feasible, but pragmatically devastating. If everyone pursues this strategy, we will have a group of solipsists who will stay caught in their own assessments, who will never come to a decision, and who will therefore fail to solve the task they are facing.

Moreover, a consensual resolution of disagreement is the only strategy which is *generalizable* to the entire group: When consensual methods are applied, the agents

[2] The epistemic tension here is a degree version of Moore's paradox. Recall that Moore's paradox consists in asserting "*P* but I don't believe that P."

typically have a hunch about each other's past performance, and it is natural to assume that none of the agents will be absolutely right. This provides an argument for conciliatory, consensual resolutions in general, not only as pragmatically superior, but also as epistemically privileged. The Lehrer–Wagner model provides us with a tractable and elegant way of successively resolving such epistemic tensions, and can, in this sense, be defended as a model of rational consensus.

4 The Non-Uniqueness Problem

The reasoning in the previous section demonstrates that there is some rational foundation to the Lehrer–Wagner model. This does not mean, however, that a particular method of eliminating the differences of opinion is the most defensible—at least as long as we do not make substantial additional assumptions. In particular, the above arguments lend support to a consensual resolution of differences of opinion, but they do not entail that not abiding by the Lehrer–Wagner model is irrational.

A similar point is made in Loewer and Laddaga (1985) and Bradley (2006). Loewer and Laddaga argue that the principle of "respecting someone's views" is not enough to justify the fact that "the correct way of taking other people's views into account is the Lehrer–Wagner procedure" (Loewer and Laddaga 1985, 84). That critique, however, falls short of addressing the theorem that Lehrer and Wagner (1981, 110) provide in order to justify the uniqueness of their approach. The point is stressed in Lehrer (1983), who defends the model as a unique rational model on formal grounds.

We argue here that it is not purely on formal grounds—that is, satisfaction of formal properties—that the uniqueness of the model ought to be defended. To make this point clearer, note that the Lehrer–Wagner model works with *relative weights* w_{ij}, which each agent i assigns to any other group member j. These relative weights are used to update the (numerical) beliefs x_i according to simple linear averaging:

$$x_i^{(2)} := \sum_{j=1}^{N} w_{ij} x_j. \tag{1}$$

But without a philosophical story of why we should update our estimates in this rather than another way, the method of updating remains arbitrary. To argue for the uniqueness of (1), we would have to assign *meaning* to the relative weights, for example by providing an account of how they are derived from past performance data. It is not clear why we should change our beliefs according to the weighted linear average (1), instead of, for instance, the weighted geometric average:

$$x_i^{(2)} := \prod_{i=1}^{N} x_j^{w_{ij}} \tag{2}$$

Allegedly, the weighted geometric average also satisfies a number of desirable mathematical properties. Thus, we are struggling to justify the uniqueness of the model on formal grounds alone, without a canonical interpretation of the relative weights w_{ij}. Natural attempts to do so, such as viewing them as success frequencies, meet substantial problems, such as accommodating the normalization of weights required by the model. As things stand, the w_{ij} are nothing more than entries in a matrix. And this prompts the question of why we should combine them according to (1) instead of according to (2), or some other model of consensus through mutual respect.

To justify a specific way of updating our estimates, we need an account of how they are produced and what they mean. For instance, Hartmann and Sprenger (2010) develop a model for elucidating the value of an unknown quantity by combining numerical estimates from different sources. The different estimates are modeled as stochastic random variables, with a given degree of precision and bias. Thus, the problem of achieving consensus is translated into the problem of combining the individual estimates optimally, that is, to find the relative weights w_i such that the linearly averaged estimator $\sum_{i=1}^{N} w_i\ x_i$ has zero bias and minimal variance. Finally, it is shown that this optimization problem has a unique solution for various families of loss functions.

Unlike Hartmann and Sprenger (2010), the Lehrer–Wagner model does not specify how estimates or weights are generated or given a specific meaning. This precludes the attempt to argue for a particular method of combining them (e.g., by linear averaging). To say that w_{ij} expresses the "part of a unit vote which i is willing to grant individual j on the basis of his respect for j" (Wagner 1978, 340) does not suffice: we need an operationalization of "proportion of a unit vote", and an account of why it is epistemically advantageous to combine these votes in a specific, linear way.

Lehrer and Wagner (1981, 138–140) give some examples where optimal relative weights can be determined by means of specific statistical modeling assumptions, akin to the model of Hartmann and Sprenger (2010). However, this route is not satisfactory for a model of convergence to consensus through mutual respect. First of all, the examples are very diverse and do not suggest a general procedure for finding, interpreting and updating respect weights. Secondly, it remains unclear why the final weights should represent a genuine *consensus* rather than an epistemically optimal aggregation procedure of numerical estimates. Thirdly, the disagreement resolution in these examples is *static* and not dynamic: on the basis of the modeling assumptions, one can directly determine the optimal weights. There is no role for an updating procedure or a combination of different forms of expertise (cf. the next section). So why should we then speak of *convergence* of opinions through mutual respect? What is the function of the matrix multiplication mechanism in the Lehrer–Wagner model?

Lehrer and Wagner fail to provide a plausible account of how an epistemically beneficial result can be retrieved by a consensual or deliberative procedure. Thus, their model cannot ground its own rationality, let alone claim epistemic superiority over other averaging methods. This vindicates a point from Sect. 2: a normatively interesting consensus model has to have a philosophical backbone for the procedures it suggests. The argument from resolving epistemic tensions, which was supposed to provide such a backbone, makes a case for conciliatory views in general, but it does not single out a unique method of consensus formation (e.g., through mutual respect). So it is questionable how we can rescue the normative force of *any* formal consensus model. Justifying a method of rational belief revision in the face of disagreement demands additional, possibly statistical, assumptions (as the Hartmann-Sprenger example makes clear); the descriptive adequacy of such assumptions will, however, only be warranted in exceptional cases.

5 The Problems with Convergence

The previous section shows how difficult it is to argue for a particular method of obtaining a rational consensus through mutual respect. This forces us to retreat a little: we cannot defend the view that it would be irrational to violate the prescriptions of, say, the Lehrer–Wagner model. Still, we can make a case for that model by investigating whether it is a psychologically *realistic* representation of rational belief revision in the light of disagreement. If this were the case, then, although being pluralists regarding which consensus models we consider *admissible*, the descriptive adequacy of the Lehrer–Wagner model would single it out among alternatives.

This section argues, however, that the idea of *gradually* resolving epistemic disagreements through mutual respect is problematic for factual disagreements on both normative and descriptive grounds. Again, we use the Lehrer–Wagner model as a case study to support our claims.

The Lehrer–Wagner model describes consensus as *convergence* of opinions: a vector of individual estimates $v = (v_1, \ldots, v_N)$ is mapped to the product $\ldots \times W_k \times W_{k-1} \times \ldots \times W_1 v$. That convergence can be interpreted in two ways: as *repeated opinion updating* (all W_k are equal to each other) where differences vanish in the limit only, or as an *aggregation of different types of information* (Kelly 2010) that correspond to different stages of the convergence process (the W_k can be different). We believe that either interpretation fails to achieve normative uniqueness, and does not convince on a descriptive level either.

Consider the temporal updating interpretation first. The group members submit anonymous position papers, and update their opinion according to the opinion of their peers, as codified in these papers (Wagner 1978). Since none of them is dogmatic, they will all move towards some form of weighted average of the individual estimates. Then this procedure is repeated until consensus is reached. But even if we put aside the problematic interpretation of the relative weights, a justification of why we iterate the procedure is difficult to find: why should we move further once we have blended our own opinion with the views of our peers? Obviously, one should not doubly count the evidence. Thus it is unclear why deliberating agents would use the same information about their peers over and over again. The only novel fact is that our peers have changed their opinions as well. But should this make any impression on us? Which information relevant to the subject matter of disagreement is contained in our peers' opinion change? Apparently, this is only a descriptive psychological statement about the degree to which they are influenced by others. It is unclear why their epistemic fluctuations *should* affect our opinion on the subject matter, as long as we are pooling opinions and not evidence. The same problems recur with respect to descriptive adequacy. It may take a very large number of iterations to reach a consensus, but in a realistic scenario, a group in deliberation is unlikely to undergo many iterations to reach a consensus. Why would deliberating agents use the same information over and over? Why wouldn't they instead change the assignment of relative weights, depending on how others revise their opinions?

These objections suggest that, if we aim at *rational* consensus, the convergence should not be interpreted in a temporal way, but rather as an aggregation of different kinds of information. We could construct a sequence of matrices $(W_k)_{k\in\mathbb{N}}$ such that each matrix codifies a specific kind of expertise: W_1 expresses belief on how competent the other agents are in the primary subject area under consideration, W_2 expresses how reliable they are at evaluating the expertise of others, etc. Second- or third-order beliefs seem to naturally affect rational belief revision: maybe some other group members know whom to trust in the question under disagreement. If you trust their judgment, you should defer to the same person they defer to. An analogous case could be made for even higher-order beliefs. Then, Lehrer and Wagner (1981) and Wagner (1978) argue that the matrices W_k need to be multiplied with each other in order to feed beliefs of different orders into the belief revision process. The matrix $W^\infty = \prod_{k=1}^{\infty} W_k$ amalgamates all relevant information, and using it for the averaging of estimates can therefore be defended as a *rational* consensus.[3]

Leaving aside the fact that the aggregative interpretation (which allows for different weight matrices) needs stronger convergence conditions than the temporal interpretation, it is still hard to see how it can succeed descriptively. In the aggregative version, the agents do not only assess the *factual* expertise of the other group members, but also their expertise at recognizing experts, at recognizing expert-recognizers, and so on. The model quickly becomes cognitively absurd, as no one would expect the agents to be able to cognitively represent, and work with, such high degrees of assessments of expertise.[4]

While there is apparently no natural interpretation to be given to the higher-order weights, they crucially affect the outcome through the mathematics of the model. This runs counter to the simple insight that our competence in assessing others becomes the smaller and the less reliable, the higher the involved orders of expertise. Thus, higher orders of expertise should in general carry less weight than

[3] Lehrer and Wagner (1981) show that allowing for different matrices in the product need not affect the convergence properties. Hence, we can neglect that case for the present discussion.

[4] Theorem 7.4 (Lehrer and Wagner, 1981, 135), shows that every weight matrix where a "chain of respect" can be established between the group members possesses a unique fixed point. Lehrer and Wagner argue that the only differential weights that can be defended as a rational consensus of the group should have this fixed point property. In particular, the consensual weights should be invariant under taking into account second-order expertise—an assumption that we find questionable. Considering only the consensual weights, thereby avoiding the iterative nature of the model, seems equivalent to assuming that there already is, in the group, a consensus on the weights needed to reach convergence of opinions. But a consensus on the relative weights is what the model is meant to produce, not to start from.

lower orders of expertise (although they should carry *some* weight)—a point that Lehrer and Wagner neglect.

On the normative side, the problem of non-uniqueness returns: the new interpretation combines information of different types by matrix multiplication. Thus, agent *i*'s respect for agent *j*, taking into account first- and second-order expertise, can be written as

$$w_{ij}^{(1,2)} = \sum_{k=1}^{N} w_{ik}^{(2)} w_{kj}^{(1)}. \tag{3}$$

In other words, *j*'s relative weight in *i*'s weight vector is the competence on the subject matter which all agents assign to *j* ($w_{kj}^{(1)}$), averaged by *i*'s judgment on these agents regarding their ability to recognize experts ($w_{ik}^{(2)}$). The non-uniquenss objection from Sect. 4 can be repeated in a different context: it is unclear why we should calculate the total weights in this particular way, except for reasons of mathematical convenience. And saying that the relative weights are assigned such that combining them in this particular way is optimal would amount to a *petitio principii*.

It is time to draw some partial conclusions. Summing up the discussion of the last three sections, we conclude that there are many objections that can be raised against a socio-epistemic understanding of the Lehrer–Wagner model. Some objections, such as the lack of descriptive adequacy and the lack of a compelling interpretation of the relative weights, are sufficiently general to lead us to question the epistemic adequacy of any consensus model based on convergence of opinions through mutual respect. We do not mean to suggest that these criticisms are decisive, but clearly they pose serious difficulties. In addition, the expert judgment literature indicates that most of the time, there is no positive correlation between epistemic performance and perceived expertise (Cooke 1991). Thus, even if the above theoretical criticisms could be addressed, the Lehrer–Wagner model might not succeed in assisting practical group decisions.

Anyone wishing to defend a model of rational consensus on factual matters must respond to each of the criticisms we outlined above. Instead of pursuing that goal here, we wish to devote the remainder of this paper to shifting the target of the Lehrer–Wagner model, and to defend it as a consensus model about non-factual matters. We contend that things are different for non-factual disagreements, and it is here that the Lehrer–Wagner model looks most promising.

6 Non-Factual Disagreements

The interpretation of the Lehrer–Wagner model as a model for resolving scientific disagreement—as representative of most factual disagreement—is standard in the literature, both for those defending the model and for those objecting to it. To provide only some examples, the assumption that the model is about factual disagreement is made in Laddaga (1977), Lehrer (1983) and Loewer and Laddaga (1985). Our criticisms above are based on this assumption, too. Such an interpretation of the model, however common, does not seem the only possible one. The title of Lehrer and Wagner (1981) is *Rational Consensus in Science and Society*, and it seems reasonable to think that disagreement in society will necessarily involve matters of value, utility, preferences, etc. Instead of thinking of disagreements about the number of jelly beans in a jar (a canonical case of epistemic disagreement), we now move on to cases where it is plausible that there is no underlying fact of the matter. As an example, consider a group of friends wishing to spend an evening out, and having to choose between going to a restaurant, a movie, or a concert. Here they must come to some agreement on matters of value, such as which of the options they ought to value the most.

In general, we now turn to cases involving questions of value, preference, priority, or aesthetic judgment.[5] It is plausible that there is no fact of the matter about whether Baroque music is better than classical music, whether Pinot Noirs are tastier than Merlots (despite the pontification in the movie “Sideways”), or whether we ought to prefer a concert to a movie.

It is worth noting that there is nothing in the Lehrer–Wagner model itself that precludes applications such as the one we are entertaining here. This is a crucial difference between the Lehrer–Wagner model and the very similar DeGroot model. The DeGroot (1974) model is explicitly about belief aggregation, whereas the Lehrer–Wagner model is more open-ended about its uses. The latter merely requires that there be disagreement among the agents about the value of a quantity, and that the agents in question assign weights to all the agents in the group. These weights express degrees of expertise in the relevant subject matter. What are we to make of this key feature of the Lehrer–Wagner model when we are entertaining non-factual disagreement? In the next sections we will take up the problem of weight assignment in the non-factual case, and provide a justification of the iterative pooling procedure for non-factual disagreements.

7 Weight Interpretation in Non-Factual Disagreements

There are at least a couple of different ways to make sense of the relative weights when the Lehrer–Wagner model is used for non-factual disagreements. First, consider the weights in terms of *coherence of values*. Here we interpret them as the degree to which the agent assigned the weight concurs with the agent assigning the weight. So w_{ij} will be a normalized measure of the extent to which agent j tracks the values of agent i.[6]

Another, and perhaps more natural interpretation of the weights is as *degrees* or *proportions of concern*. For example, an agent might care about the opinions of

[5] If you’re a cognitivist about ethical or aesthetical values, replace our examples with your own favorite cases where you take it that there is no fact of the matter, or with preference aggregation.

[6] The Regan metric (Regan et al. 2006) for weights of respect (which derives the weights matrix from the distance between the various combinations of quantity values) lends itself very naturally to this interpretation. But there is also room for alternative metrics: for some examples, see Martini (2012). An agent can assign less than maximal respect to herself; she might, for instance, have less than perfect confidence in her ability to know what she wants or what is best for her.

some group members more than about others. This care can take many forms but it need not reflect power structures, and can be based on mutual closeness. Personal or family relationships seem like a poor basis for assigning weights in the epistemic case, but they are a natural basis for reaching consensus about non-factual matters. For instance, when going out for dinner with a keynote speaker, it is reasonable to give her special weight regarding the choice of the evening program. Moreover, an ethics of care licenses such differential treatment, with special treatment to those closest to you (Gilligan 1982; Held 2005; Slote 2007).

This interpretation is different from the previous one offered. The degree of concern interpretation makes no judgment about how far the views of others are from your own. You might acknowledge that some disagree with you more than others but still, you weigh the former more heavily than the latter because you are concerned to satisfy them. Both these interpretations of the weights apply to the kind of non-cognitivist disagreement we are interested in in this section.

To our mind, the second interpretation is particularly promising. It can be spelled out further in terms of *respect for mutual preferences*. An agent might prefer Merlot over Pinot Noir, but at the same time have a desire to conform to peer group preferences, to be considered as more cultured (perhaps as a result of watching "Sideways", and taking the type of wine snobbery portrayed in that movie too seriously), not to offend others, or for other reasons. So the agent distributes weights of respect in such a way as to reflect her desire to concur with the other group members. For example, if our Merlot-loving agent has some desire to conform to her peers, she will assign high weights to other group members as well, although they could turn out to prefer Pinot Noir over Merlot. Thus, we could interpret the relative weights as the extent to which the preferences of our agent depend on the preferences of the rest of the group.

This move solves the first and most salient problem raised earlier—finding a coherent and natural interpretation of the mutual respect weights—by stipulation. It also gives us a strategy to tackle the problem of non-uniqueness: if the relative weights indicate the *proportion* to which an agent *prefers* to agree with someone else, the linear averaging method of Eq. (1) is indeed rational. In the factual case, the rationality of the Lehrer–Wagner model depends on the link between the relative weights and the available information on the other group members' epistemic competence and past performance. Without such a link, the relative weights are ostensibly meaningless, and repeated pooling is an arbitrary procedure.

Compare this to the non-factual case. The relative weights can be understood as describing the extent to which others' preferences affect our own preferences and the extent to which the desire to concur with others changes our views. This interpretation is close to French's (1956) original scenario where the relative weights denote social impact, and it gives us an interesting model of preference change via role models and peer influence.[7]

[7] This strikes us as a very plausible formal model of at least some preference change and is thus worthy of further investigation. See Becker (1996), Bradley (2007, 2009) and Pollak (1970, 1978) for more on the fascinating issues associated with preference change.

Thus, in contrast to the epistemic case, the linear averaging method naturally flows from the interpretation of the relative weights. It should also be noted that linear averaging can be shown to be the unique method that satisfies a number of properties which are commonly accepted in social choice theory (see Dietrich and List 2011), one of the fundamental fields dealing with group aggregation problems. This does not yet show that linear averaging is *the* paramount method for aggregating preferences, but it gives a good rationale for using linear averaging rather than other methods.

To repeat, the reason why it does not seem rational to adopt the same strategy for the epistemic case is that, in that case, the rationality of the outcome has to be defended in terms of factual accuracy. To explicate what it *means* to reasonably average with someone else's opinion, we would have to presuppose a formal model in which such a compromising attitude is shown to foster factual accuracy. As we have argued in Sects. 4 and 5, Lehrer and Wagner fail to provide such a model in defense of (repeated) linear averaging. By contrast, as we have argued in this section, no such justification is required in the non-factual case because factual accuracy is not a criterion for assessing the outcome.

8 Iterated Pooling in Non-Factual Disagreements

The repetition of the pooling procedure with the same weights can also be defended more easily in the non-factual case. In the epistemic case, rational disagreement after a round of updating had to be explained by informational differences, and it was not clear why the updating procedure should be repeated. In the non-factual case, the agents' desire to concur with the rest of the group does not vanish when disagreement persists. Reaching a consensus can be presupposed as a joint goal that trumps persisting differences in the group, as it is plausibly the case for our dinner party, or a committee that needs to come a unanimous decision. Repeating the procedure is a very natural way to satisfy that joint desire, and to resolve the tension between other-regarding and egoistic preferences.

It might be objected that the *effective weight* which the Lehrer–Wagner model, after N repetitions, assigns to one of our peers will typically not agree with the degree of concern that we assigned to her beforehand. For example, an agent may reserve most of the weight to herself so that she will pull the group into her direction and receive a high effective weight. It then seems that we have agreed to a result that we are unhappy with, raising the question of whether we are still facing a *consensus* on non-factual matters, rather than a covert dictatorship.

This objection is certainly valid in the epistemic case: First, theoretical considerations and empirical literature (e.g., Armstrong 2001) suggest that combining diverse opinions is epistemically advantageous for reducing bias and enhancing performance. Second, the only way to rationally justify giving greater consideration to a stubborn peer is based on the—questionable—assumption that he or she is much more competent than ourselves on the matter at hand.

Contrast this with your respect for others in a non-factual case. Going back to the case of having dinner with a friend, assume that you assign equal weights. This respect is a product of our social relationship, and independent of dietary requirements, or views about what food is best. But if your friend persists in her preference for a Thai restaurant, your reaction will differ from the epistemic case. There, it was necessary to ascribe to her a high second-order expertise in order to justify another round of averaging. Whereas, in the non-factual case, repeating the averaging procedure can be more easily defended: you *learn* that her preferences are stronger than yours, and take this information into account. Perhaps she is a vegetarian, perhaps she loves Thai food and rarely gets the opportunity to eat it, perhaps she has had traumatic experiences with the Dutch restaurant that you suggested as an alternative. Unlike in the factual case, the fact that an agent may reserve most of the weight to herself does not prompt the problem that individual (epistemic) bias is multiplied in the group decision.

Thus, even if you are concerned for all group members to the same degree, you may rationally give greater consideration to a friend who has stronger opinions than yourself. By discovering the *strength* of her preferences, which is independent of your respect for her, you become more inclined to give such greater consideration, and to accept a result that is close to her original suggestion. The resulting consensus can be regarded as an equilibrium state that is operationalized by a repeated averaging procedure, without explicit elicitation of strength of preferences.

It should be noted, however, that stubbornness need not be due to strong preferences, but can also be the result of *indifference* to the preferences of others. If that were the case, then it appears as if there would be no reason for you to give greater consideration to your friend. But is this phenomenological difference relevant for the rationality of the chosen procedure? We believe it is not, and illustrate our view with an example.

A vegan whose diet 'requires' that her friends choose a specific restaurant may, in one perspective, be described as just having a very strong preference on the subject matter (that we should account for), whereas another perspective might depict her as unwilling to make even a slight dietary compromise to accommodate her friends. This difference is, however, purely perspectival—the two descriptions are empirically indistinguishable. We accommodate our vegan friend because she is our friend and because we would like to respect her preferences. As long as no conflict with fundamental personal values exists, respect for our friends' preferences seems to be independent of either the strength or nature of those preferences.

The Lehrer–Wagner model also implies that, in the end, the group members do not only agree on the issue in question, but also on their effective weights. While this is not very plausible in the epistemic case, it is a natural result of the gradual discovery of the strength of the individual preferences in the non-factual case. We agree to go to the Thai restaurant because we have *discovered* that our friend is much less indifferent to the choice of the restaurant than we are.

To conclude this section, the lines of thought in this and the previous section suggest a coherent interpretation of the Lehrer–Wagner model. It is an interpretation that does not rely on there being any underlying facts of the matter. A rational consensus is reached through mutual respect. Weights are assigned and interpreted consistently, there is a justification for the iterative nature of the model, and the practical conflicts between perceived expertise and epistemic performance do not arise. The Lehrer–Wagner model may thus well survive as a useful consensus model, and give sufficient answers to the criticisms advanced earlier in this paper. For instance, Regan et al. (2006) use the Lehrer–Wagner model for a disagreement over an environmental resource allocation dispute. These considerations do not only rationally ground the Lehrer–Wagner model in the non-factual case, they are, *a fortiori*, also good arguments for resolving non-factual disagreements through respect-based opinion convergence.

Finally, we should note there are disagreements about issues where factual and value issues are mixed. Take an academic job search. There might be some objective facts about who the very best job candidates are and who the worst are. It is often the case, however, that disagreements about who the single best candidate is arises from value judgments about the research interests of the candidate, and even the inherent worth of his or her areas of specialization. When groups are deliberating about such issues, factual and non-factual considerations become entangled. A formal epistemologist on the committee might assign higher weight to a logician's opinion than a ethicist's opinion. The Lehrer–Wagner model might be a useful tool for achieving consensus in such cases—that is, whenever there is at least an element of non-factual disagreement. Nevertheless, any analysis of these "mixed" disagreements should preferably disentangle the factual and non-factual components (or to give good reasons for not doing so). Of course, this will, in general, be difficult in practice.

9 Conclusion

In a variety of circumstances where groups are charged with a decision, consensual resolutions are required, some times for legal reasons, or in order to ensure smooth collaboration in further stages of common work. In this paper, we have investigated the rationality of procedures that reach convergence of opinions based on mutual respect, with a particular focus on the Lehrer–Wagner model.

In epistemic contexts, whether a group adopts a final judgment (e.g. a group average) as a consensus, or as a mere compromise, depends on whether there has been any genuine belief revision by the individuals in the group. Thus, a formal consensus model has to be motivated and interpreted so that there is a plausible story about why the process of arriving at an agreement might be understood as a *rational belief revision process*.

The best-known and most significant representative of that class of models is the Lehrer–Wagner model, but we have outlined a number of objections to applying the model to settle epistemic disputes. We highlighted some poorly-motivated features of the model and note the lack of a convincing account of why it should be thought of as a (uniquely) rational way of changing beliefs in the face of peer disagreement.

We went on to show, however, that the machinery of the Lehrer–Wagner model is suitable for resolving non-factual disagreements, such as disagreements on aesthetic matters, ethical questions, or individual preferences. We argued that the objections to the epistemic interpretation of the model do not carry over to this new interpretation. On the contrary, we are able to arrive at a consistent and intuitive justification for resolving non-factual disagreements through mutual respect, namely as a gradual discovery of mutual strength of preferences. If the Lehrer–Wagner model aspires to be a model of rational consensus, non-factual disagreement is a promising arena for it to operate.

Finally, in the same sense that epistemic consensus models are rivalled by sophisticated voting and averaging procedures, non-factual consensus models are rivalled by preference aggregation models from social choice and bargaining theory. It is therefore important to stress once more that a rational defense of the Lehrer–Wagner model presupposes the desirability of a genuine consensus. Under these conditions it demonstrates how mutual respect can drive a rational consensus on non-factual matters.

Acknowledgments We would like to thank Rachel Briggs, Filip Buekens, and three anonymous referees for useful comments on earlier drafts of this paper. This work was initiated by two working groups held at the Tilburg Center for Logic and Philosophy of Science and the University of Sydney, both funded by a Dutch Science Foundation Internationalization Grant ("Formal Modeling in Social Epistemology") and by the Australian Centre of Excellence for Risk Analysis. Jan Sprenger expresses his thanks to the Netherlands Organization for Scientific Research (NWO) for their support of his work through a Veni grant (No. 016.104.079).

References

Armstrong, J. S. (2001). Combining forecasts.' In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer.

Becker, G. S. (1996). *Accounting for tastes*. Cambridge, MA: Harvard University Press.

Bradley, R. (2006). Taking advantage of difference of opinion. *Episteme: A Journal of Social Epistemology*, *3*, 141–155.

Bradley, R. (2007). The kinematics of belief and desire. *Synthese*, *156*(3), 513–535.

Bradley, R. (2009). Becker's thesis and three models of preference change. *Politics, Philosophy and Economics*, *8*(2), 223–242.

Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, *116*, 187–217.

Cooke, R. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford: Oxford University Press.

DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*, 118–121.

Dietrich, F., & Christian List. (2011). The aggregation of propositional attitudes: Towards a general theory. *Oxford Studies in Epistemology, 3*, 215–234.

Elga, A. (2007) Reflection and disagreement. *Noûs*, *41*, 478–502.

French, J. R. P., Jr. (1956). A formal theory of social power. *Psychological Review*, *63*, 181–194.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*, Cambridge, MA: Harvard University Press.

Hartmann, S., & Sprenger, J. (2010). The weight of competence under a realistic loss function. *The Logic Journal of the IGPL*, *18*, 346–352.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation (JASSS), 5*(3). http://jasss.soc.surrey.ac.uk/5/3/2.html.

Held, V. (2005). *The ethics of care: Personal, political, and global*. Oxford: Oxford University Press.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389.

Jackson, M. O., & Golub, B. (2007). Naïve learning in social networks: Convergence, influence and wisdom of crowds. *Fondazione Eni Enrico Mattei Working Papers* (2007).

Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne & T. Szabo (Eds.), *Oxford studies in epistemology* (Vol. 1, pp. 167–196). Oxford: Oxford University Press.

Kelly, T. (2010). Peer disagreement and higher order evidence. In R. Feldman & T. Warfield (Eds.), *Disagreement*. Oxford: Oxford University Press.

Laddaga, R. (1977). Lehrer and the consensus proposal. *Synthese*, *36*, 473–477.

Keith, L. (1976). When rational disagreement is impossible. *Noûs*, *10*, 327–332.

Keith, L. (1983). Rationality as weighted averaging. *Synthese*, *57*, 283–295.

Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society*. Dordrecht: Reidel.

Loewer, B., & Laddaga, R. (1985). Destroying the consensus. *Synthese*, *62*(1), 79–95.

Martini, C. (2012). Consensus formation in networked groups. In W. R. de Henk, H. Stephan & S. Okasha (Eds.), *EPSA philosophy of science: Amsterdam 2009* (pp. 199–215). Dordrecht: Springer.

Page, S. E. (2007). *The difference*. Princeton: Princeton University Press.

Pollak, R. A. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy*, *78*(4), 745–763.

Pollak, R. A. (1978). Endogenous tastes in demand and welfare analysis. *American Economic Review*, *68*(2), 374–379.

Regan, H., Mark, C., & Lisa, M.-N. (2006). A formal model for consensus and negotiation in environmental management. *Journal of Environmental Management*, *80*, 167–176.

Slote, M. (2007). *The ethics of care and empathy*. Routledge, London.

Steele K., Helen R., Mark C., Burgman, M. A. (2007). Right decisions or happy decision-makers?. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, *21*, 349–368.

Wagner, C. (1978). Consensus through respect: A model of rational group decision-making. *Philosophical Studies*, *34*, 335–349.