

4-23-2019

Completely Theorized Agreements. A Different Reading of the Consensus Paradox Hypotheses

Marta Wojciechowska

London School of Economics and Political Science, m.wojciechowska@lse.ac.uk

Follow this and additional works at: <https://www.publicdeliberation.net/jpd>

Part of the [Other Political Science Commons](#), and the [Political Theory Commons](#)

Recommended Citation

Wojciechowska, Marta (2019) "Completely Theorized Agreements. A Different Reading of the Consensus Paradox Hypotheses," *Journal of Public Deliberation*: Vol. 15 : Iss. 1 , Article 14.

Available at: <https://www.publicdeliberation.net/jpd/vol15/iss1/art14>

This Symposium is brought to you for free and open access by Public Deliberation. It has been accepted for inclusion in Journal of Public Deliberation by an authorized editor of Public Deliberation.

Completely Theorized Agreements. A Different Reading of the Consensus Paradox Hypotheses

Abstract

This article contributes to the debate on *the consensus and deliberation*. While the relevant literature claims that consensus undermines further deliberation, this article argues that it depends on the aim of the process. In particular, I argue that if the aim of deliberation is understood as reaching a certain epistemic level, reaching consensus does not need to decrease the rationality of the group. In short, such deliberation is a process of debate, reason-giving and listening which aims at establishing a result of certain epistemic value. In order to shed new light on the debates on the consequences of consensus for further deliberation, I introduce a detailed conceptualization of a full agreement. I call it *Completely Theorized Agreements*. In this article, I argue that reaching consensus in an epistemic setting does not need to have negative consequences. Further, I argue, that the truth-tracking quality of deliberation need not be worse in a group that reached a full consensus as opposed to a partial one.

Author Biography

Marta Wojciechowska is a Fellow at LSE100 and a PhD Candidate at the London School of Economics.

Keywords

Democracy; Deliberation; Consensus; Truth

Acknowledgements

This article benefited from comments of participants of ‘Deliberation after consensus: Democracy, epistemic quality and public discourse workshop’ (2014) and the LSE Winter Workshop (2015). In particular, I am grateful to Henrik Friberg-Fernros, Johan Karlsson Schaffer, H el ene Landemore, Martin Ebeling, Tom Parr, Ellie Knott, JPD Editors and two anonymous reviewers for their helpful comments.

There is a paradox within deliberative theory. The early deliberative literature treats consensus as the ultimate aim of rational debate (Cohen, 1996; Habermas, 1996; Miller, 1992). While the current literature has moved away from the notion of an ideal agreement, some form of consensus remains a core part of the deliberative project. Consensus is present either as an actual agreement on rules governing society (Bohman & Richardson, 2009), an ideal aim (Knops, 2007), or it takes a form of a partial or meta-agreement (Curato, Dryzek, Ercan, Hendriks, & Niemeyer, 2017; List, 2002; Niemeyer & Dryzek, 2007). The paradox arises as deliberative literature also argues that at least minimal disagreement is necessary for fruitful deliberation (Thompson, 2008) and a lack of diverse opinions within the deliberative forum can lead to ‘group think’ (Solomon, 2006), polarization (Mercier & Landmore, 2012; Sunstein, 2002), or the hindering of collective action (Warren, 2017). Furthermore, when new facts appear, previously reached agreements may impinge on the rationality of further deliberation (Friberg-Fernros & Schaffer, 2014).

In this paper, I argue that reaching consensus need not have negative effects for further deliberation if a deliberation has an epistemic aim. In short, such epistemic deliberation is a process of debate, reason-giving and listening, which aims at establishing ‘truth’ or ‘correctness’ (Estlund & Landmore, 2018, p. 113; Fuerstein, 2014, p. 285; List & Goodin, 2001, p. 277; Min & Wong, 2018, p. 3). Such deliberation is present in the scientific process, among juries within the U.S. judiciary system and members of truth and reconciliation commissions. In a case of epistemic deliberation, the negative consequences of reaching consensus for further deliberations are potentially more troubling. More precisely, if the previously established consensus were to undermine the rationality of the revision, the social trust placed in these institutions would also be undermined.

In order to explore the consequences of reaching consensus for further deliberation in a public setting with an epistemic aim, I focus on *the consensus paradox hypotheses* introduced by Friberg-Fernros and Schaffer (2014). At present, the hypotheses offer the most detailed prediction of negative consequences of reaching consensus. In short, the hypotheses predict that the more the deliberation reaches consensus, the worse will be the quality of the deliberation thereafter. Consequently, the hypotheses expect the paradox to become more pronounced as the deliberating group reaches full consensus. In order to shed new light on the debates on the consequences of consensus for further deliberation, I introduce a detailed conceptualization of a full agreement. I call it *Completely Theorized Agreements* (hereafter CTAs/CTA), which refer to existing literature on consensus (in particular: Elster, 1998a; Sunstein, 1995). In accordance with Sunstein’s logic, full consensus occurs when participants agree on the individual choices and on the

general principles behind their choices. CTA is, therefore, a special case of a consensus and enables comparison between full and partial consensus. In this article, I argue that reaching consensus in an epistemic setting does not need to have negative consequences. Further, I argue that the truth-tracking quality of deliberation need not be worse in a group that reached a full consensus as opposed to a partial one. While this argument does not challenge the *consensus paradox hypotheses per se*, as they are an empirical conjuncture, it does represent a development of the existing literature on the characteristics of an epistemic consensus in deliberation (e.g., Beatty & Moore, 2010; Dryzek & Niemeyer, 2006, 2010; Estlund & Landmore, 2018; Fuerstein, 2014; Jezierska, 2019; Mercier & Landmore, 2012). As such, this is an optimistic argument that is of special importance to deliberations within the institutions of social trust.

The Consensus Paradox in an Epistemic Setting

What is the essence of the phenomenon known within the deliberative literature as the *consensus paradox* (Friberg-Fernros & Schaffer, 2014)? The paradox arises from the contradictory aims and requirements of deliberation posed by the core literature. The early deliberative literature claims that the principal merit of deliberation lies in its ability to produce a substantive, rational consensus and unanimity of preferences (Cohen, 1996; Elster, 1986; Habermas, 1996; Miller, 1992). However, empirical studies show that unanimous results are very difficult to achieve during collective decision-making (List, 2007; List, Luskin, Fishkin, & McLean, 2012), and that, even if they were possible, reaching consensus is undesirable (Mouffe, 2000a, 2000b). Newer approaches within the literature claim that valuable deliberation can end with some form of compromise, i.e., incompletely theorized agreements (see Mansbridge et al., 2010).¹ Despite this, some authors claim that even if an ideal consensus is difficult to achieve, deliberative participants should either treat it as a potential aim (Knops, 2007) or opt for some form of partial agreement (Curato, Dryzek, Ercan, Hendriks, & Niemeyer, 2017; List, 2002; Niemeyer & Dryzek, 2007).

However, even when a full agreement is treated hypothetically, or can be realized to some degree, a tension persists between consensus and the pluralism of values. When the group reaches a full consensus, participants are no longer in a position to deliberate, because at least minimal disagreement is essential to any fruitful deliberation (Thomson, 2008). Furthermore, as J. S. Mill (1859/2011) famously

¹ Incompletely theorized agreement happens when participants of deliberation agree on the same outcome but for different reasons (Sunstein, 1995).

argues, any opinions, if not discussed, “will be held as a dead dogma, not a living truth” (p. 35). Similar opinions held by the participants of the deliberative forum can undermine the process, and result in overconfidence (Mercier & Landemore, 2012), groupthink (Baron, 2005) or polarization (Sunstein, 2002). In practice, consensus, once reached, is expected to restrict the deliberative capacity of a group by limiting its pluralism (Friberg-Fernros & Schaffer, 2014).

The consensus paradox notes that it may be fruitless to deliberate after having reached full agreement. In particular, the *consensus paradox hypotheses* offer, currently, the most precise scholarly claims in respect of the consequences of consensus. The hypotheses claim that consensus, once established, negatively influences the quality of deliberation thereafter (Friberg-Fernros & Schaffer, 2014). The *consensus paradox hypotheses* appear in weaker and stronger forms. The weaker version (H1) states that “the more an agreement on the policy issue approximates consensus, the less subsequent public discourse on that issue will improve rationality” (Friberg-Fernros & Schaffer, 2014, p. 108). The stronger version (H2) states that “the more an agreement on a policy issue approximates consensus, the less rational subsequent public discourse will be on that issue” (Friberg-Fernros & Schaffer, 2014, p. 108).²

However, depending on how rationality is defined, it is possible to interpret the hypotheses differently. Friberg-Fernros and Shaffer (2014, p. 109) show that rationality takes on varied definitions within deliberative research. In their article, they refer to this concept as outcome rationality; in other words, the “epistemic quality of the collective decisions or judgements resulting from deliberation” (see Cooke, 2000, p. 952).³ The authors operationalize outcome rationality in terms of the internal validity of the arguments used during deliberation. In particular, they concentrate on the number and quality of reasons presented during the discussion as a measurement of increased or decreased rationality (Friberg-Fernros &

² The weak *consensus paradox hypothesis* claims that consensus, once established, does not obstruct the group’s ability to deliberate, but limits the ability of the process to increase collective rationality. Friberg-Fernros and Shaffer (2014) expect that reaching a consensus will eliminate incentives for participants to develop new arguments and question existing ones, at least for some time. The strong hypothesis is more ambitious. It states that “the more an agreement on a policy issue approximates consensus, the less rational subsequent public discourse will be on that issue” (Friberg-Fernros & Schaffer, 2014, p. 108). In other words, the claim is that to the degree that deliberation reaches a consensus, the rationality of a group will be subsequently diminished. Thus, the strong hypothesis claims that a particular group reaching a consensus decreases the group’s rationality. Friberg-Fernros and Shaffer (2014) argue that this hypothesis can be confirmed when, after having reached a consensus, members of the group would forget their initial arguments.

³ For other ways to define rationality, see Bohman (1998), Goodin (2000), and Pettit (2001).

Schaffer, 2014, pp. 109-111). In this interpretation of rationality, the central claim of the *consensus paradox hypotheses* can be rephrased as: the more a group reaches consensus during deliberation, the worse will be the ability of its members to present well-supported arguments thereafter.

However, here I examine the consensus paradox from a different perspective. More precisely, I interpret rationality as a tool for reaching a result of epistemic value (Fuerstein, 2014, p. 285). This view assumes that there is an independent standard with which to evaluate the outcomes of political decisions and procedures like ‘truth’ or ‘correctness’.⁴ This interpretation of rationality is more coherent with the epistemic arguments within the deliberative literature. The ability to track correct decisions is one of the core premises defended by proponents of deliberation (see Cooke, 2000). Similarly, Neblo (2015) writes about an epistemic warrant of deliberation: “just as poker is no longer poker if we are not trying to win, rational discussion is no longer rational discussion if we are not trying to find the right answer” (p. 106).

This premise, or warrant, is not only applied to claims that deliberation produces better decisions; it often also forms the basis for deliberations in scientific environments. Such scientific deliberations have vital public importance. Furthermore, under certain conditions, it is possible for members of the public to retrace them (Gutmann & Thompson, 2004; Moore, 2018). As such, scientific deliberations are public.⁵ However, it is also possible to find the epistemic warrant in strictly political deliberations. For example, the assumption that the members of the U.S. Supreme Court can reach the correct decision during deliberation is the essence of the practices of judicial review and of democracy in the U.S. Similarly, the aim of truth and reconciliation commissions in post-conflict societies is to establish the truth about past events, thereby contributing to the process of national healing.⁶

⁴ Landemore (2013) calls this view political cognitivism. However, the debate on the existence of this standard is much older (see Cohen, 2009). My point here is not to take a stance on either side of this debate, but merely to show that it is possible to interpret rationality as referring to an epistemic standard.

⁵ For the opposite view, see Elster (1998b).

⁶ As Rotberg and Thompson (2000) note, for post-war societies, “in order to come fully to terms with their brutal pasts, they must uncover, in precise detail, who did what to whom, and why, and under whose orders. They must seek, at least, thus to uncover the truth – insofar as this aim is humanly and situationally possible after the fact” (p. 3). The most famous example of such a commission is the Truth and Reconciliation Commission in post-apartheid South Africa; however, there are other examples of such commissions all over the world: e.g., Rwanda, Argentina, Chile, Guatemala and Uganda (Sarkin, 1999).

My aim here is to show that, contrary to the scope of Friberg-Fernros & Schaffer's argument (2014), the consensus paradox can be investigated from the perspective of the ability of the deliberative group to track the truth. From this perspective, it is possible to rephrase the *consensus paradox hypotheses*. The revised version of a weaker hypothesis (H1') would state that the more an agreement on a policy issue approximates consensus, the less subsequent public discourse on that issue will improve the ability of the group to track correct decisions (Friberg-Fernros & Schaffer, 2014, p. 108). The revised version of the stronger hypothesis (H2') would now state that the more an agreement on a policy issue approximates consensus, the ability of the group to track the correct decision will decrease (Friberg-Fernros & Schaffer, 2014, p. 108).

The consensus paradox is potentially more troubling if the deliberation's epistemic qualities are understood as truth-tracking, rather than if it is operationalized as the quantity and quality of arguments used during the deliberation. This is because, as discussed above, many of the arguments for deliberation, as well as those more generally in support of democracy, are based on the assumption that deliberation can reach a result of epistemic value. The paradox can be especially problematic for deliberative groups in which wider democratic literature sees an important source of epistemically good decisions, for example among scientists, juries, or truth and reconciliation committees.

To analyze how the *consensus paradox hypotheses* work, according to this interpretation, in the next section I introduce Completely Theorized Agreements as an ideal type of a full consensus. Following these hypotheses, it could be expected that after reaching a CTA, the rationality of the group will decrease more so than if the group had reached a partial agreement. However, I argue that this need not be the case.

Completely Theorized Agreements in Epistemic Deliberations

Completely Theorized Agreements refer to Sunstein's and Elster's line of argument (Elster, 1998a; Sunstein, 1995).⁷ In accordance with Sunstein's logic, full consensus can occur when participants agree on the individual choices and on the general principles behind their choices. In other words, full agreement includes consensus on the following: values [normative consensus]; beliefs about how

⁷ In the literature, it is also possible to find this kind of agreement labelled as "universal consensus" (Dryzek & Niemeyer, 2010, p. 94; Femia, 1996, p. 368).

particular actions will map those values [epistemic consensus]; and, finally, actions which need to be taken [preference consensus] (Elster, 1998a, p. 100). CTA is consistent with the current literature on full consensus and its theoretical requirements (Dryzek & Niemeyer, 2010; Femia, 1996) and complements the literature on partial consensus.⁸

Despite the fact that such full consensus is difficult to achieve (List, 2007; List et al., 2012), this article treats it as an empirical possibility. This type of agreement could occur when participants agree on the individual choices and on the general principles behind those choices (after Sunstein, 1995). To achieve full agreement, members of the group need to match their convictions, about what should be done, with their beliefs, concerning how those convictions could be realized (Elster, 1998a). Based on these elements, Dryzek and Niemeyer (2010) recognize three different kinds of agreements. Normative consensus relates to the agreement on values and about what should be done in a certain situation. Epistemic consensus concerns a situation when participants agree on how particular choices match certain normative values. Finally, preference consensus is achieved when participants agree on actions which should be taken (Dryzek & Niemeyer, 2010). While there are many other ways of conceptualizing agreements, I treat CTA as a useful way of conceptualizing theoretical requirements for full consensus since they emphasize the precise partial agreements.⁹ In line with the *consensus paradox hypotheses*, I investigate what would need to happen for the singular deliberation to result in a full agreement. Thus, I investigate what influence a full agreement could have on the rationality of a group after a process of deliberation. CTA could appear when participants agree on normative values, when they have the same beliefs about how those values can be realized and, finally, when their actual choices are unanimous. As such, it is a demanding agreement.

To illustrate the requirements for reaching CTA, I consider a hypothetical decision-making procedure within an academic environment. Let us imagine that at one of London's universities, a previously used room becomes empty. The management staff of the university, academic staff representatives and student representatives gather to discuss the room's future purpose. During the first discussion, it appears that these three groups diverge from each other in terms of their norms, beliefs and preferences. The management staff of the university holds the view that any changes should benefit the school financially. To this end, they believe that renting

⁸ For example, see Bohman (1996, p. 89), Dryzek & Niemeyer (2006), and Mansbridge et al. (2010).

⁹ Other examples of conceptualizing consensus include Moore & O'Doherty (2014), Mansbridge et al. (2010), and Thompson (2008).

out the room would be beneficial. As a result, they want to rent out the room for commercial use. The academic staff holds the view that a university's most important value is research, meaning that enhancing research should be the principal aim of the university's endeavors. They believe that keeping the room will be beneficial as it can be used as an additional research space and, therefore, they want to keep the room. Finally, there are student representatives who are concerned about the costs of studying in London. Their main value is student finance, and they aim to decrease the costs for students. They believe that keeping the room is beneficial as it can be transformed into a student lounge in which students can heat up their own meals and not spend so much money on high-street lunches. As a result, they opt to keep the room. The division of preferences, beliefs and values among these three groups is illustrated in Table 1.¹⁰

	VALUE:	BELIEF:	PREFERENCE:
<i>Management</i>	Finance	Renting the room is beneficial	Rent the room
<i>Academic Staff</i>	Research	Keeping the room is beneficial	Keep the room for research
<i>Students</i>	Finance	Keeping the room is beneficial	Keep the room and use it as a common room

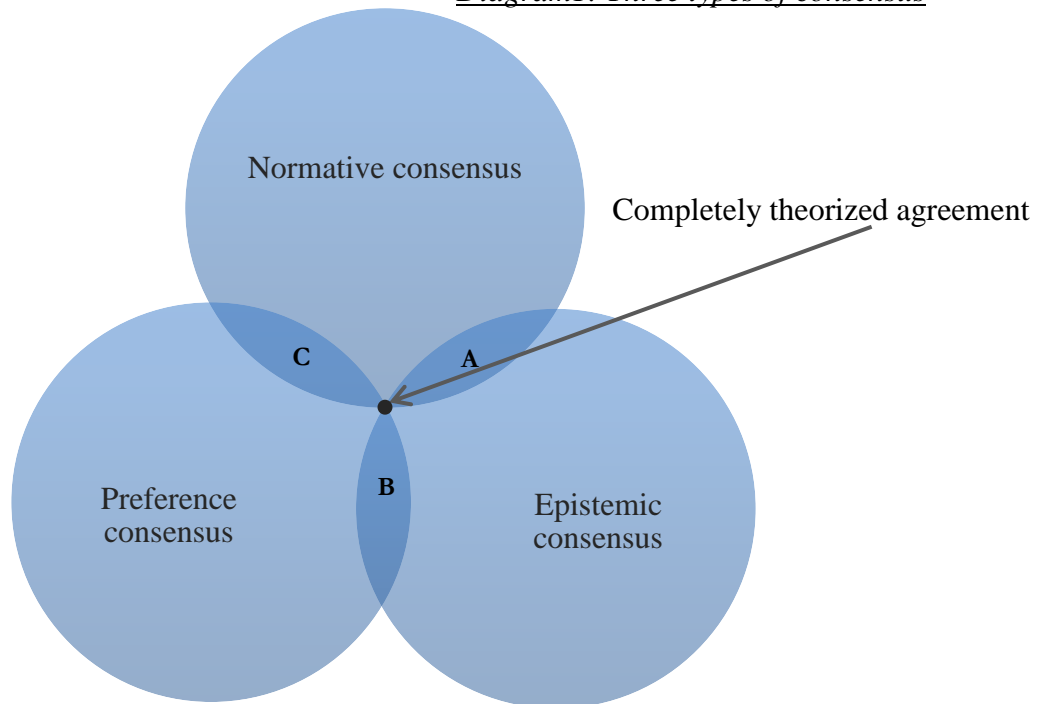
Table 1: Empty Room Example. Division of Values, Beliefs and Preferences

At the beginning of the decision-making process, these three groups have different norms, beliefs and preferences. However, it is possible to imagine that, as the deliberation proceeds, some participants may change their norms, beliefs and preferences. In other words, it is possible that they will achieve a partial consensus. Diagram 1 below illustrates examples of the types of possible partial consensus within these three groups.¹¹

¹⁰ The logic of the university example is similar to the one given by Elster (1998a), as well as Dryzek and Niemeyer (2010, p. 94). However, the aim of this paragraph is not to discover an entirely new case of divisions of values, beliefs and preferences, but instead to illustrate possible convergence towards partial consensus.

¹¹ As a way of conceptualizing analytical problems, this diagram follows the lead of List (2011).

Diagram 1: Three types of consensus



It is possible that as a result of deliberation, in the process of providing reasons that other members find convincing, academic staff and students agree that research is the university's core value. It is agreed during deliberations that the room ought to be used as an additional research space. Both students and academic staff believe that keeping the room is beneficial. However, even if those groups agree on the general principle that should direct the decisions of the university, and they both hold the view that keeping the room is beneficial, they can still differ in their actual preferences. For example, the academic staff would like to keep the room and transform it into an additional research room for staff, while students may like to keep the room to be transformed into a research room for the use of students. This partial consensus is illustrated by sphere A above, in which there is normative and epistemic consensus, but still disagreement in terms of preferences.

The second type of partial consensus may occur when participants have similar beliefs about the result of certain actions, and have similar preferences, but disagree on values. This partial consensus is reflected by sphere B in the diagram. This partial agreement can be illustrated by a situation in which, after deliberation, the university's management could convince academic staff that renting the room to an NGO will be beneficial to all groups involved. Both groups would be able then to

decide to rent out the room. However, the management thinks that renting is good for the university's financial security, while academics hope to increase their impact by having the NGO across the corridor. Therefore, both groups agree that renting out the room is beneficial, and they decide to do so; however, the reasons why they rent out the room are different. Furthermore, the norms underlining those reasons are different. This situation reflects an incompletely theorized agreement in which participants agree on certain decisions, but they do not agree on the reasons or higher-order norms for this decision (Sunstein, 1995). This kind of partial agreement often occurs when participants have not completely theorized the basis of their (partial) agreement (Sunstein, 1995).

Finally, it is also possible that during deliberation participants will agree on general values and particular actions, but not on how certain decisions reflect those values. This would be possible if the management convinced the student body that renting out the room will increase the general situation of the university, which, as a consequence, will also be felt by students. Here, management could provide examples that confirm how a better financial situation has also benefited students, since in those cases there was more money to be spent on student facilities and bursaries. Students agree that, in general, in order to secure more money for everyone, the best course of action would be to rent the room. In this situation, both the management and students hold the view that finance is the most important variable to consider. Both groups agree to rent the room. However, students have agreed to it despite the belief that the best way to realize their own value (finance) is to keep the room. This partial consensus is based on a normative and preference agreement, while lacking an agreement on beliefs. This partial agreement is illustrated by sphere C in the diagram.

In the 'empty room' example, a completely theorized agreement will be difficult but not impossible to achieve. It could only happen when all three groups hold the same norms, the same beliefs about how those norms are reflected by certain choices and, finally, have the same preferences.¹² Consider how a completely theorized agreement could be established during a one-off deliberation event. First, this kind of agreement would require a normative consensus. In the 'empty room' example, a change in values (finance vs. research) was shown to be quite probable. However, this is because it concerned lower-order values. A switch in higher-order

¹² All of these partial agreements are necessary conditions for CTA. For example, in a situation of a perfect normative and epistemic consensus, there may still be different choices that would be coherent with the identical norm and beliefs about how this norm matches different choices. Similarly, the preference and epistemic consensuses together do not guarantee CTA. This is because various norms can lead to the same epistemic beliefs and preferences regarding the choices given.

normative values, such as moral values, is more difficult. Higher-order values are the result of socialization and are rooted in one's understanding of the world. A change in higher-order values is more likely as a result of more prolonged education or indoctrination; however, it is unlikely to happen during a one-off deliberation event. Therefore, a normative consensus on higher-order values can be established during deliberation, yet it is only probable among people who already hold the same higher-order values prior to the event. That said, such a situation can happen when people deliberate to find the best solution, for example during scientific deliberation in which the values reflect the common aim of finding the result best supported by evidence.

Establishing an epistemic consensus requires a similarity of beliefs before the deliberative process commences. Following Mackie (2006), I assume that beliefs do not exist on their own, but rather comprise a complex network of beliefs that exert a mutual influence either positively or negatively. Furthermore, as Chambers (1996) points out, a change in beliefs is gradual and fragmented. This means that it takes time to change one's beliefs. As beliefs are influenced by one another and are not prone to rapid change, a deep epistemic consensus during a one-off deliberation event can only be achieved if participants hold similar beliefs before deliberating.

In addition to a normative and epistemic consensus, CTA also requires agreement on actual preferences. Following Landwehr (2005), I assume that preferences are connected to normative values and, as a result, are considerably stable. In this interpretation, a change of preferences is based on two factors, namely hearing new arguments and some sort of utility maximization. Preferences in this model are not solely the representation of beliefs. Landwehr (2005) argues that preferences gradually decrease in stability. This means that a change of preferences is easier for lower-order preferences or when higher-order preferences are similar. Therefore, to follow the Landwehr model, a deep preference consensus can be achieved if members of the group hold considerably similar preferences prior to the one-off deliberation event.

One-off deliberation can result in CTA if members hold similar norms, beliefs and preferences prior to the deliberation. But this can also occur if members are able to engage in an in-depth analysis in respect of their choices, beliefs and values. In order to achieve this type of consensus, participants need to discuss their reasons for supporting particular options, as well as analyze why they believe certain things are true and why they support given norms. Finally, they also need to agree.

To summarize, CTA is a special case of a consensus, one that is demanding to achieve. In the next section, I will illustrate the usefulness of this concept for

investigating the premises and consequences of claims concerning rationality in deliberation by analyzing the consensus paradox. However, CTA, as an ideal type, can also be employed when investigating other claims on consensus within the field of deliberative democracy. While CTA is a more technical term for a full consensus, the term clarifies the core conditions for achieving such consensus. Hence, while full consensus can be simply interpreted as a full agreement on the preferences of a whole deliberating group, CTA flags up the demanding normative and epistemic conditions of a full consensus. As a result, it is a precise analytical tool.

Completely Theorized Agreements and the Paradox

CTA enables us to shed new light on consensus paradox hypotheses in an epistemic setting. To recall, the quality of epistemic deliberation refers to the ability of the group to reach a good epistemic outcome. This epistemic standard refers to an external standard, like truth or correctness (see Estlund & Landemore, 2018; Min & Wong, 2018). Here, the quality of the deliberation does not depend on the quality of the arguments per se; rather, it derives from the ability of the group to identify the correct result. However, most likely, the quality of arguments facilitates the identification of the truth. Due to this external metric of the quality of deliberation in an epistemic setting, reaching consensus need not have negative consequences for the epistemic quality of deliberation. Moreover, the truth-tracking quality of deliberation need not be worse in a group that reached a full consensus as opposed to a group that reached a partial agreement. To illustrate these points, in this section I consider a hypothetical case of epistemic deliberation in a scientific, laboratory environment.

For the sake of argument, imagine a hypothetical deliberation among scientists concerning the result of the laboratory experiment.¹³ The hypothetical laboratory group tests the responsiveness of animal cells to a new drug that they have developed. Let us assume that the group has discussed the results and comes to a full consensus. They agree that the drug they have tested, in light of the evidence, is effective. Reaching CTA in a laboratory group is possible if the group shares the same values [normative consensus].¹⁴ In order to reach full consensus, scientists would have to be convinced that the evidence is conclusive [epistemic consensus] and that the evidence is relevant to the theory tested [preference consensus]. CTA does not occur if the group is pressured into arriving at the consensus or prefers to

¹³ For a detailed anthropological analysis of scientific deliberations see Latour and Woolgar (1986).

¹⁴ Here, such shared values would relate to the norm that the aim of the scientific endeavor is to find the scientific truth.

reach an agreement on the basis of other factors, like time-pressure or funding competition.

Now imagine that the drug has been used in a trial, but the results were negative. The initial scientific group would probably engage in further laboratory work to repeat the experiments. The group could deliberate further on the interpretation of the new results. Here, the laboratory group may either support the previous interpretation or retract it. The possible outcomes of the initial and consequent deliberations are illustrated in Table 2 below.

Scenario:	Outcome of the First Deliberation:	Outcome of the Consequent Deliberation:	Change in the Epistemic Value:
1	Incorrect	No change: still incorrect	No change
2	Incorrect	Conclusion retracted	Increased
3	Correct	No change: still correct	No change
4	Correct	Conclusion retracted	Decreased

Table 2: Matrix of Possible Outcomes in the Laboratory Deliberation

Following the consensus paradox hypotheses, the epistemic quality of the second deliberation should either be constant or decreased. However, the epistemic quality of such scientific deliberation, at least in principle, need not decrease (see Table 2). Furthermore, even if forgetfulness and conformism were to take place after reaching CTA (see Friberg-Fernros & Schaffer, 2014), this would not necessarily decrease the epistemic quality of the outcome. This is because the quality of decision-making in an epistemic public deliberation relies on the ability of group members to reach the correct outcome, not on their ability to provide more and better supported arguments. In the example above, the scientists could have provided very limited arguments, but this would not have determined the correctness of the outcome. In epistemic deliberation, the epistemic quality of the deliberation is not a property of the process. That said, the good arguments have an instrumental value and can help to secure a good epistemic outcome.

It is also possible that the scientific group will move from the correct interpretation and arrive at the incorrect one (see Table 2). However, if repeated deliberation would not at least *sometimes* bring about a correction of the results, then scientific progress would cease to be possible.¹⁵ Repeated deliberation in scientific laboratory

¹⁵ For the argument to hold, it is enough that science progresses at least some of the time.

environments, at least sometimes, needs to produce a correction of the initial result in order for science to progress.

It is notable that in the scientific deliberation, if the rationality is defined as better illustrated by the evidence, the epistemic quality may not be affected even if the group does not reach CTA but instead obtains some sort of partial agreement. However, the possibility of reaching CTA is important for a comparative aspect of the consensus paradox hypotheses. To recap, the consensus paradox hypotheses, in the revised version, states that the more the deliberative group reaches consensus, the more the ability of the group to track correct decisions will not increase (H1') or will even decrease (H2') as a result of reaching consensus. However, in the epistemic deliberation in which the 'correct' answer exists and the deliberative quality relates to the ability to track this answer, the fact of prior full (as opposed to partial) consensus is inconsequential. In the epistemic setting, the quality of deliberation is a property of reaching a particular result; it is not dependent on any previous characteristics of a group, such as full or partial sameness of views. Rather, it is more likely for the opposite case to be correct. Partial consensus can signify the presence of other factors in reaching the agreement, like time pressures or demands to achieve results. These additional pressures may negatively influence the group's ability to discover the correct decision (see Fuerstein, 2014, p. 285). As Estlund and Landmore (2018) note, in "problem-solving contexts, consensus as an ideal outcome of deliberation retains an epistemic appeal as a 'marker' of truth, signaling that no one knows or can construct a better idea" (p. 125).

Conclusion and Limitations

In this article, I have offered a contribution to the debate on *consensus* and *deliberation*. While the literature concerned with the paradox agrees that consensus undermines further deliberation, this article argues that this need not be the case. In particular, I have argued that if the aim of deliberation is understood as attaining a certain epistemic level, then reaching consensus does not need to decrease the rationality of the group further. To this end, I have analyzed *consensus paradox hypotheses* and, by introducing a hypothetical scenario, investigated whether the paradox still holds if rationality is understood as an ability to track a correct decision in public deliberation. I further introduced an example of full consensus, namely CTA. This type of consensus includes a deeply theorized normative, epistemic and preference consensus. This type of agreement is difficult yet not impossible to achieve; indeed, one example can be found in the form of a result of scientific deliberation (see Estlund & Landmore, 2018; Fuerstein, 2014). This type of consensus enables us to investigate the consequences of full consensus for the epistemic result and demonstrates that the truth-tracking quality of deliberation

need not be worse in a group that reached a full consensus than in one that reached a partial consensus.

As a final point, I would like to address the limitations of this article. While the argument does not challenge the *consensus paradox hypotheses per se*, as they are an empirical conjuncture, it does develop the literature on characteristics of an epistemic consensus in public deliberation. The reason why this article does not engage empirically with the consequences of reaching consensus for further deliberation is the character of the deliberative setting that it investigates. It is possible to empirically investigate the rationality of deliberation after reaching consensus if the result is operationalized by an internal validity of the arguments used during deliberation (Friberg-Fernros & Schaffer, 2014, pp. 110-111). If the rationality is operationalized as a better epistemic outcome, it is possible to investigate increased or decreased rationality by referring to an external metric of 'truth' or 'correctness.' However, during the scientific deliberations, such truth is often discovered during the process through laboratory experiments, and such investigation is beyond the scope of the field of social sciences to which this article belongs.

Another possible objection concerns the focus of the article. The main focal point here has been the consensus paradox, which is an important yet considerably narrow discussion within the literature. However, a wider aim of this article has been to address epistemic deliberation and consensus within the public setting. The importance of such deliberations lies in the social trust put into institutions that seek epistemic consensus. This, in turn, contributes to broader debates on the aims of truth and trust in democratic societies. Similarly, the introduction of the CTA enables a full conceptualization of the requirements of full consensus, which will prove useful to future projects on the topic of consensus. As such, this article hopes to encourage further studies on the issues in question.

References

- Baron, R. S. (2005). So right it's wrong: Groupthink and the ubiquitous nature of polarized group decision making. *Advances in Experimental Social Psychology*, 37, 219-253.
- Beatty, J., & Moore, A. (2010). Should we aim for consensus? *Episteme*, 7, 198-214. doi:10.3366/E1742360010000948
- Bohman, J. (1996). *Public deliberation: Pluralism, complexity, and democracy*. Cambridge, MA: MIT Press.
- Bohman, J. (1998). Survey article: The coming of age of deliberative democracy. *Journal of Political Philosophy*, 6, 400-425. doi:10.1111/1467-9760.00061
- Bohman, J., & Richardson, H. S. (2009). Liberalism, deliberative democracy, and “reasons that all can accept.” *Journal of Political Philosophy*, 17, 253-274. doi:10.1111/j.1467-9760.2008.00330.x
- Chambers, S. (1996). *Reasonable democracy: Jürgen Habermas and the politics of discourse*. Ithaca, NY: Cornell University Press.
- Cohen, J. (1996). Procedure and substance in deliberative democracy. In S. Benhabib (Ed.), *Democracy and difference. Contesting the boundaries of the political* (pp. 93-119). Princeton, NJ: Princeton University Press.
- Cohen, J. (2009). Truth and public reason. *Philosophy & Public Affairs*, 37, 2-42. doi:10.1111/j.1088-4963.2008.01144.x
- Cooke, M. (2000). Five arguments for deliberative democracy. *Political Studies*, 48, 947-969. doi:10.1111/1467-9248.00289
- Curato, N., Dryzek, J. S., Ercan, S., Hendriks, C. M., & Niemeyer, S. (2017). Twelve key findings in deliberative democracy research. *Daedalus*, 146(3), 28–38. doi:10.1162/DAED_a_00444
- Dryzek, J. S., & Niemeyer, S. (2006). Reconciling pluralism and consensus as political ideals. *American Journal of Political Science*, 50, 634-649. doi:10.1111/j.1540-5907.2006.00206.x
- Dryzek, J. S., & Niemeyer, S. (2010). *Foundations and frontiers of deliberative governance*. Oxford, UK: Oxford University Press.
- Elster, J. (1986). The market and the forum. In J. Elster & A. Hylland (Eds.), *Foundations of social choice theory* (pp. 103-132). Cambridge, UK: Cambridge University Press.

- Elster, J. (1998a). Deliberation and constitution making. In J. Elster (Ed.), *Deliberative democracy* (pp. 97-122). Cambridge, UK: Cambridge University Press.
- Elster, J. (1998b). Introduction. In J. Elster (Ed.), *Deliberative democracy* (pp. 1-18). Cambridge, UK: Cambridge University Press.
- Estlund, D., & Landmore, H. (2018). The epistemic value of democratic deliberation. In A. Bächtiger, J. S. Dryzek, J. Mansbridge & M. Warren (Eds.), *The Oxford handbook of deliberative democracy* (pp. 113-131). Oxford, UK: Oxford University Press.
- Femia, J. (1996). Complexity and deliberative democracy. *Inquiry*, 39, 359-397. doi:10.1080/00201749608602427
- Friberg-Fernros, H., & Schaffer, J. K. (2014). The consensus paradox: Does deliberative agreement impede rational discourse? *Political Studies*, 62, 99-116. doi:10.1111/1467-9248.12103
- Fuerstein, M. (2014). Democratic consensus as an essential byproduct. *Journal of Political Philosophy*, 22, 282-301. doi:10.1111/jopp.12040
- Goodin, R. E. (2000). Democratic deliberation within. *Philosophy & Public Affairs*, 29, 81-109. doi:10.1111/j.1088-4963.2000.00081.x
- Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton, NJ: Princeton University Press.
- Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. Oxford, UK: Polity Press.
- Jeziarska, K. (2019). With Habermas against Habermas. The bifurcation between consensus and deliberation in Habermas's theory of democracy. *Journal of Public Deliberation*.
- Knops, A. (2007). Debate: Agonism as deliberation - On Mouffe's theory of democracy. *Journal of Political Philosophy*, 15, 115-126. doi:10.1111/j.1467-9760.2007.00267.x
- Landmore, H. (2013). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton, NJ: Princeton University Press.
- Landwehr, C. (2005). Rational choice, deliberative democracy, and preference transformation. *Studies in Social and Political Thought*, 11, 40-68. doi:10.1.1.335.8785
- Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2nd ed.). Princeton, NJ: Princeton University Press.

- List, C. (2002). Two concepts of agreement. *The Good Society*, 11, 72-79.
doi:10.1353/gso.2002.0011
- List, C. (2007). Deliberation and agreement. In S. W. Rosenberg (Ed.), *Can the people govern? Deliberation, participation and democracy* (pp. 64-81). Basingstoke, UK: Palgrave Macmillan.
- List, C. (2011). The logical space of democracy. *Philosophy & Public Affairs*, 39, 262-297. doi:10.1111/j.1088-4963.2011.01206.x
- List, C., & Goodin, R. E. (2001). Epistemic democracy: Generalizing the condorcet jury theorem. *Journal of Political Philosophy*, 9, 277-306.
doi:10.1111/1467-9760.00128
- List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2012). Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. *The Journal of Politics*, 75, 80-95.
doi:10.1017/S0022381612000886
- Mackie, G. (2006). Does democratic deliberation change minds? *Politics Philosophy Economics*, 5, 279-303. doi:10.1177/1470594X06068301
- Mansbridge, J., Bohman, J., Chambers, S., Estlund, D., Føllesdal, A., Fung, A., . . . Martí, J. L. (2010). The place of self-interest and the role of power in deliberative democracy. *Journal of Political Philosophy*, 18, 64-100.
doi:10.1111/j.1467-9760.2009.00344.x
- Mercier, H., & Landmore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, 33, 243-258. doi:10.1111/j.1467-9221.2012.00873.x
- Mill, J. S. (1859/2011). *On liberty*. Luton, UK: Andrews UK Limited.
- Miller, D. (1992). Deliberative democracy and social choice. *Political Studies*, 40, 54-67. doi:10.1111/j.1467-9248.1992.tb01812.x
- Min, J. B., & Wong, J. K. (2018). Epistemic approaches to deliberative democracy. *Philosophy Compass*, 13(6). doi:10.1111/phc3.12497
- Moore, A. (2018). Deliberative democracy and science. In A. Bächtiger, J. S. Dryzek, J. Mansbridge & M. Warren (Eds.), *The Oxford handbook of deliberative democracy* (pp. 640-654). Oxford, UK: Oxford University Press.
- Moore, A., & O'Doherty, K. (2014). Deliberative voting: Clarifying consent in a consensus process. *Journal of Political Philosophy*, 22, 302-319.
doi:10.1111/jopp.12028

- Mouffe, C. (2000a). Deliberative democracy or agonistic pluralism. Reihe Politikwissenschaft / Political Science Series 72.
- Mouffe, C. (2000b). For an agonistic model of democracy. In N. O'Sullivan (Ed.), *Political theory in transition* (pp. 113-130). New York, NY: Routledge.
- Neblo, M. A. (2015). *Deliberative democracy between theory and practice*. Cambridge, UK: Cambridge University Press
- Niemeyer, S., & Dryzek, J. S. (2007). The ends of deliberation: Meta-consensus and inter-subjective rationality as ideal outcomes. *Swiss Political Science Review*, 13, 497-526. doi:10.1002/j.1662-6370.2007.tb00087.x
- Pettit, P. (2001). Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11, 268-299. doi:10.1111/0029-4624.35.s1.11
- Rotberg, R. I., & Thompson, D. (Eds.). (2000). *Truth v. justice: The morality of truth commissions*. Princeton, NJ: Princeton University Press.
- Sarkin, J. (1999). The necessity and challenges of establishing a truth and reconciliation commission in Rwanda. *Human Rights Quarterly*, 21, 767-823. doi:10.1353/hrq.1999.0042
- Solomon, M. (2006). Groupthink vs the wisdom of crowds: The social epistemology of deliberation and dissent. *The Southern Journal of Philosophy*, 44, 28-42. doi:10.1111/j.2041-6962.2006.tb00028.x
- Sunstein, C. R. (1995). Incompletely theorized agreements. *Harvard Law Review*, 108, 1733-1772. doi:10.2307/1341816
- Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy*, 10, 175-195. doi:10.1111/1467-9760.00148
- Thompson, D. (2008). Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, 11, 497-520. doi:10.1146/annurev.polisci.11.081306.070555
- Warren, M. E. (2017). A problem-based approach to democratic theory. *American Political Science Review*, 111, 39-53. doi:10.1017/S0003055416000605