Beyond Habermasian Discourse Ethics

The Conceptualization of a New Feasibility Framework

Stijn Koenraads
Beyond Habermasian Discourse Ethics:
The Conceptualization of a New Feasibility Framework

Stijn Koenraads

Master thesis in Political Theory
Department of Political Science
Nijmegen School of Management
Radboud University Nijmegen

Student number 4024664
Words (references excluded) 34,714
Supervisor Dr Marco Verschoor
Date of completion 22 June 2014
ABSTRACT

Habermasian discourse ethics is a widely known deliberative democratic theory, often praised for its ideals. Nevertheless, as several authors have pointed out, it seems to be a theory that is not feasible, or at least not feasible enough. A theory that is infeasible may contain valuable ideals, but is not suited to practical reality. Exploring a new way in which Habermas's theory can be made more feasible is the aim of this thesis. By pointing out the weaknesses of Habermasian theory, both in general and concerning feasibility issues specifically, it is argued that the theory's feasibility must be increased. Based on existing literature concerning feasibility, a conception of feasibility is formulated. Through the application of this conception to Habermasian discourse ethics, the theory's infeasibility is ascertained. Furthermore, it is argued that even the conception developed thus far is not entirely accurate nor complete, since feasibility considerations should go beyond the aspects of feasibility that other authors discuss. The impact of parts of a theory as well as the feasibility of the parts of a theory themselves should be considered as well, it is argued. For this purpose, a new feasibility framework is developed, in which the impact of parts of a theory, the feasibility of the parts of a theory as well as the feasibility of the preferred outcome of a theory have a place. This new framework is applied to an important part of Habermas's discourse ethics: the discourse rules. The new feasibility framework provides the opportunity to judge a theory, or even merely parts of it, on its feasibility in a more complete and accurate sense than authors have done so far. The framework cannot only be utilized to judge and enhance the feasibility of Habermas's theory, but can be used in the same way for other (political) theories as well.
PREFACE

During the process of writing this master’s thesis and completing the master’s programme in political science as a whole, I not only learnt a lot about political science and research methods, but I also discovered and explored various qualities, weaknesses and personality traits that I have. The programme has taught me a great deal and I want to thank the people that have directly contributed to this.

First and foremost, I would like to sincerely thank my thesis supervisor, Dr Marco Verschoor, for providing me with his valuable and helpful comments as well as his knowledge of the topic of this thesis. You have invested a part of your time in me and my work, and I can only hope that the results of my (and your) efforts correspond to your inspiring guidance. Furthermore, many thanks to Dr Thomas Eimer, Prof. Dr Marcel Wissenburg and Dr Andrej Zaslove for using their expert knowledge to fill my head with valuable thoughts, whilst always giving me the liberty to think, feel and write in my own way.

Moreover, I wish to thank Annemarie Verschragen, Jaap van Hoogmoed, Remco Castelein, Roos van Straaten and Wietske Degen for their efforts to provide me with useful comments from their own ever-illuminating perspectives. You are friends and colleagues who are critical, honest and interested in my work. Your comments have positively influenced the writing process and its result.

Also, my parents, family and friends have given me the opportunity to (re)discover myself and my academic interests as well as fulfil the duties of becoming a beginning scholar in political theory. I am grateful to you.

Finally, I wish to thank you, Silvana, for always being there for me, providing me with useful comments as well as precious emotional support. You have given me the time and affection I needed. The process of writing this thesis would not have been the same without you.

Stijn Koenraads

Nijmegen, 21 June 2014
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>III</td>
</tr>
<tr>
<td>PREFACE</td>
<td>IV</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 THE MORAL THEORY OF DISCOURSE ETHICS</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Habermas’s Moral Theory of Discourse</td>
<td>5</td>
</tr>
<tr>
<td>2.1.1 The rules of discourse</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Habermas’s Critics</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1 Consensus and procedure</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2 Problems with the principle of universalization (U)</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3 Impartiality and generalized versus concrete others</td>
<td>14</td>
</tr>
<tr>
<td>2.3 The desirability of a deliberative democratic theory</td>
<td>19</td>
</tr>
<tr>
<td>3 A BIPARTITE CONCEPTION OF FEASIBILITY</td>
<td>22</td>
</tr>
<tr>
<td>3.1 The need for a feasible discourse ethics</td>
<td>22</td>
</tr>
<tr>
<td>3.2 Important features of feasibility</td>
<td>25</td>
</tr>
<tr>
<td>3.3 An account of feasibility</td>
<td>28</td>
</tr>
<tr>
<td>4 THE FEASIBILITY OF HABERMASIAN DISCOURSE ETHICS</td>
<td>34</td>
</tr>
<tr>
<td>4.1 Hard constraints concerning discourse ethics</td>
<td>34</td>
</tr>
<tr>
<td>4.1.1 Momentum and inevitability</td>
<td>34</td>
</tr>
<tr>
<td>4.1.2 The violation of physical laws (nomological consistency)</td>
<td>36</td>
</tr>
<tr>
<td>4.1.3 A historically fixed starting point</td>
<td>37</td>
</tr>
<tr>
<td>4.1.4 Logical consistency</td>
<td>37</td>
</tr>
<tr>
<td>4.1.5 Biological/psychological and natural human ability constraints</td>
<td>39</td>
</tr>
<tr>
<td>4.1.6 Stability</td>
<td>41</td>
</tr>
<tr>
<td>4.2 Soft constraints concerning discourse ethics</td>
<td>42</td>
</tr>
<tr>
<td>4.2.1 Economic, cultural, institutional and motivational constraints</td>
<td>42</td>
</tr>
<tr>
<td>4.2.2 Natural human abilities</td>
<td>47</td>
</tr>
<tr>
<td>4.2.3 Stability</td>
<td>48</td>
</tr>
<tr>
<td>4.3 Constraints for discourse rules specifically</td>
<td>48</td>
</tr>
<tr>
<td>4.3.1 The logical level</td>
<td>48</td>
</tr>
<tr>
<td>4.3.2 The procedural level</td>
<td>49</td>
</tr>
<tr>
<td>4.3.3 The process level</td>
<td>50</td>
</tr>
<tr>
<td>4.4 Discourse ethics and feasibility</td>
<td>52</td>
</tr>
<tr>
<td>5 A NEW FEASIBILITY FRAMEWORK</td>
<td>53</td>
</tr>
<tr>
<td>5.1 A broader notion of feasibility</td>
<td>53</td>
</tr>
<tr>
<td>5.2 A new set of discourse rules</td>
<td>56</td>
</tr>
<tr>
<td>5.2.1 The impact of the discourse rules</td>
<td>57</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Since the late 1980s, Jürgen Habermas's notion of “discourse ethics” has been prominent in political theory (cf. Habermas, 1990, 1993, 1996a). In short, discourse ethics connects communicational logic to morality and proposes a deliberative approach to moral theory. In this approach, the validity of moral norms is judged through an actual discourse that follows certain discourse rules to prevent the results of this discourse from being unjustifiable to anyone who is affected by its outcomes. These rules prescribe that such a discourse must be characterized by sincerity, inclusiveness, communicative equality and absence of coercion (Habermas, 1990, pp. 87-89). If we act as if we were in an “ideal speech situation”, Habermas argues, we should not want these rules to be violated. In order to reach a rationally motivated consensus that can be accepted by all people affected, discourse participants must abide to these rules, which they would rationally want to hold for the discourse (Habermas, 1973; 1990, pp. 82-86, 103-104).

As opposed to the monological theories of, for instance, John Rawls (1999a), Charles Taylor (1994) and Axel Honneth (Honneth, 1995; Fraser & Honneth, 2003), discourse ethics is a moral theory with an emphasis on practice: it is a dialogical moral theory and it prescribes that actual discourses be organized (Habermas, 1990, pp. 65-66, 82, 93-94; Rummens, 2007b), instead of, for instance, justifying moral norms in a monological, individual and hypothetical manner, like in Rawls’s original position (Rawls, 1999a). This emphasis on praxis generates a requirement that discourse ethics has to fulfil: because it is a theory that claims to be practical and requires actual discourses to take place, it should be applicable. For what is the value of a theory that emphasizes praxis but is not applicable? (cf. Günther, 1993) Furthermore, the dialogical relation between ideal and non-ideal theory demands that we at least consider the feasibility of a strongly ideal theory and vice versa (cf. Mutz, 2008, pp. 523-534; Valentini, 2012). Feasibility considerations serve as a “check” for theories that aim at developing grand ideals without paying too much attention to their practicability. Therefore, it is important to investigate the feasibility of political theories. Conversely, moral or ethical ideals prevent us from merely being guided by efficiency or feasibility concerns, because they show us what is right or good for us to do.

The question Habermas’s focus on praxis raises is to what extent we can use Habermasian discourse ethics to justify (and/or apply) moral norms in reality. This is very important for Habermas, since a critical theory like his claims to reflect the needs and norms of

---

society, but to also offer us practical possibilities for enlightening and emancipating citizens (cf. Geuss, 1981).

The feasibility of discourse ethics has been investigated from several angles. Günther (1993), for instance, has argued that only one type of discourse is not enough. Günther asserts that norms should not only be *justified*, but also *applied*. Therefore, we need two separate discourses: one for justification and one for application. The first discourse should be organized to justify norms and reach a general consensus; the second one to discuss the appropriateness of those norms for specific cases. This way, the norm is not merely justified without actually applying it. This would make norms both universal and effective in particular situations. Nevertheless, there is still a practicability problem, which Habermas also recognizes (1996b, p. 1518; 1998a, pp. 365-369): a rational consensus is impossible in reality, both empirically and conceptually. Consensus cannot be the goal of discourse ethics, for that would mean that whenever a consensus is reached, discourse would bring itself to a halt. For discourse is not needed anymore when a consensus has been reached (cf. Rummens, 2012).

Sanders (1997) has also criticized discourse ethics. In short, she thinks deliberation itself should not be striven after, because it leaves power structures and the dominance of certain groups intact. Moreover, it distracts us from other problems, like problems concerning inclusion or recognition – which she thinks are more basic. Social hierarchies and oppression on the basis of race, gender or otherwise seem to make participation in a deliberative discourse alienating and subject to the power structures that exist in society. Sanders supports these arguments with empirical data.

Chantal Mouffe (2005) also offers a critique on Habermas's theory. She contends that Habermas is wrong in stating that the political relates to democracy in an antithetical way. Mouffe believes that “[the] emphasis on the ever present possibility of the friend/enemy distinction and the conflictual nature of politics constitutes the necessary starting point for envisaging the aims of democratic politics” (Mouffe, 2005, pp. 13-14). In other words, she believes that conflict is necessarily an element of democracy, and defines democratic politics.

Mutz (2008) criticizes deliberative democratic theories from an empirical standpoint. She asserts that the elements they contain are often impossible to measure (and exist) together, since some elements of deliberative theories contradict others in those same theories. In order to make such theories measurable, she believes we should design several “middle-range” theories that contain different elements of deliberative theories, instead of try to analyse and discuss those theories as “grand theories”.

Nevertheless, I believe the notion of feasibility these authors focus on is too thin. A first problem is that the impact of different elements of a theory is rarely discussed when feasibility considerations are addressed. In my opinion, the relevance of such elements for the (outcome of
the) theory is an important – though often neglected – part of feasibility. A second problem is
that feasibility considerations often restrict their focus to the outcome of a theory itself. The
feasibility of the separate elements of the theory under discussion is rarely discussed, which also
makes feasibility considerations incomplete. In this thesis, I will develop a new feasibility
framework that accounts for these two problems and grasps the different elements of feasibility
in a more complete and more accurate sense than previous conceptions of feasibility. I will argue
that Habermasian discourse ethics is infeasible. I will not only investigate the feasibility of the
most important aspects of this theory, but I will also propose a way in which its feasibility can be
increased. Through applying the new feasibility framework I will develop, one part of
Habermas's theory will be altered and made more feasible, namely the part concerning the
discourse rules. I will do so by developing a ranking of the discourse rules. This feasibility
framework does not merely provide us with a manner in which the feasibility of Habermasian
discourse ethics can be increased, but also present us with a new way of conceptualizing
feasibility in general. Furthermore, this framework can be applied to many other (deliberative)
political theories, since it is not specifically designed for Habermas’s theory. By considering the
practical implications of political theories in a new way, this is also relevant for scholars or
others who try to implement or apply political theories to social reality.

I must point out that I will not develop a new basis for the same theory. Rather, my aim
will be to present a new conception of feasibility, judge the most important parts of Habermas’s
theory on the basis of this conception, embed the conception of feasibility in a broader feasibility
framework and propose a way in which discourse ethics can be made more feasible. In doing
this, I will answer the following question:

"How can Habermasian discourse ethics be made more feasible?"

I will answer the above-mentioned research question in a number of steps. First, I will introduce
Habermas's discourse ethics, and some important points of critique on it, as well as discuss why
a theory of deliberative democracy should be investigated in the first place (chapter 2).
Subsequently, I will explain why a political theory, and Habermasian discourse ethics
specifically, should be feasible, at least to a certain extent. Also, a variant of the conception of
feasibility often used in the literature will be presented (chapter 3). After this, I will explore to
what extent Habermasian discourse ethics is feasible by applying the conception of feasibility
developed in chapter 3 to the most important elements of his theory as well as the discourse
rules specifically (chapter 4). In chapter 5, I will argue that the conception of feasibility used
thus far is incomplete by discussing the elements of feasibility that are no part of this account.
The conception used thus far is a good start, I will assert, but fails to account for several other
important elements of feasibility. I will embed the notion of feasibility developed thus far in a broader framework of feasibility that not only makes the conception of feasibility more complete, but can be applied to research fields other than political science or political theory as well. Also, this framework will be applied to a specific part of Habermasian discourse ethics, namely the discourse rules. Furthermore, I will propose a way in which the feasibility of the discourse rules, and thence the (path to the) preferred outcome of Habermas's theory, can be enhanced. Finally, a concluding discussion will be presented (chapter 6).
2 THE MORAL THEORY OF DISCOURSE ETHICS

What follows now is a theoretical discussion of Habermas's discourse ethics. Subsequently, the most prominent points of critique on his theory will be touched upon in order to spell out in detail various aspects that are central to his theory and that are best explained through addressing these points of critique. Finally, it will be discussed why a theory of deliberative democracy is desirable in the first place.

2.1 Habermas's moral theory of discourse

One of the notions that strongly relate to the modern concept of 'deliberative democracy' is Habermas's idea of “discourse ethics” (Habermas, 1990; 1993). Partially developed through Habermas's views on communication and speech (as elaborated in his Theory of Communicative Action (1984/1987) and other works), discourse ethics approaches moral theory in a way that links the logic of communication and argumentation to morality. Habermas argues that every (speech) act is based on the implicit agreement on every factor that is relevant to the situation of the act. The validity of the norms that are agreed upon is determined by a moral discourse that holds the idea of an “ideal speech situation” and has certain discourse rules that should be maintained to establish a communicatively rational discourse (Habermas, 1990, pp. 57, 65, 87-89, 93). Communicatively rational actions are different from instrumentally (or strategically) rational actions in that they aim at reaching a reasonable consensus and understanding rather than a practical compromise. Instrumentally rational actions are aimed at success in a practical sense and expect people to act merely on the basis of their self-interests (Habermas, 1998b; Rummens, 2007a, pp. 18-20). Habermas argues that people are, in fact, able to think beyond their own self-interests and able to reach, through reasonable discussion with other citizens, an agreement on what their common interests are and how these can, and should, play a role in decision-making and laws. This way, the laws and policies that are based on the common interest form a boundary for instrumental action: a person is permitted to make as many decisions that serve her self-interest as she wants, as long as the common interest is not endangered by this. Importantly, Habermas assumes that people can cooperate and be loyal towards laws or policies that are in the common interest, and not only try to fulfil their self-interests (Rummens, 2007a, pp. 21-24). People do not try to get the best out of it by competing with each other in a kind of market, but talk with each other about what the common good is in a

_________________________________________________________________________

2 Direct critique on the feasibility of discourse ethics will be discussed in chapters 4 and 5.
kind of forum\(^3\) (cf. Elster, 1986). I will elaborate on the benefits of Habermas's account over the instrumentally rational account later.

Habermas's social theory is concerned with maintaining and creating social order (Finlayson, 2005, p. 47). In his view, social order is based on meaning and validity. A society is not only held together by the shared understanding, reasons and meanings that are developed in the *lifeworld*, but also by the instrumental actions in the *system*. The lifeworld and the system are the two forms of social being Habermas acknowledges. In the lifeworld, communicative action and discourse – a reflective form of speech that aims at reaching a rationally motivated consensus (Habermas, 1984, p. 42) – take place; in the system, instrumental actions take place. Habermas believes that people in contemporary societies often live in a society that is not in their interests. This is because people’s lifeworld is *colonized* by parts of the system, such as capitalist markets and bureaucratic administrations. This means that the lifeworld is restrained. Instrumental and rational action takes the place of (most) communicative action and, thus, people are led astray from their own, ultimate goals. A loss of autonomy and meaning in their lives is the consequence.

Because of this, Habermas argues, we must preserve the lifeworld and must not let it be dominated by the system (Habermas, 1987; Finlayson, 2005, pp. 51-57). In order to achieve this, moral order is needed: people must be able to adhere to demonstrably valid norms, which in turn determine whether a certain action is right or wrong. A wrong action is prohibited by a valid moral norm, a right one is permitted by it. In order to determine whether a norm is valid or not, it is tested for its universalizability: does the norm entail a universal interest for all people that are possibly affected by it in their capacity as participants in a practical discourse? (compare Habermas’s “discourse principle” (D) below) This is to be found out by subjecting the norm to a moral discourse, in which it will become clear, through rational communication, whether the norm can be accepted or not. If it can, a rational agreement in such a discourse can be attained; if not, the norm is not seen as valid.

Habermas believes that interaction can only be successful if the facts, norms and expectations that are relevant in a certain situation are agreed upon. To make this agreement possible, a principle of universalization (U) is needed (Habermas, 1990, p. 57). If we want to develop norms that are valid, we must have a certain argumentation for it. This argumentation must be substantial at least in some way, for arguments that are “compelling in terms of logical inference […] reveal nothing substantively new” (ibid., p. 63). However, because substantial arguments always depend on an individual interpretation, they cannot offer an ultimate

---

\(^3\)This is often found in empirical studies investigating the so-called “ultimatum game”: even if individuals have the opportunity to disadvantage the individuals they negotiate with and maximize their own profits, negotiation often results in relatively equal gains for both individuals. Also see Sulkin & Simon, 2001.
argumentation for validity of norms either. To bridge this gap, we need the principle of universalization (ibid., p. 65):

(U) All affected can accept the consequences and side effects [a norm's] general observance can be anticipated to have for the satisfaction of everyone’s interests (and these consequences are preferred to those of known alternative possibilities for regulation).

Habermas justifies this moral principle by explicating that “every argumentation, regardless of the context in which it occurs, rests on pragmatic presuppositions from whose propositional content the principle of [universalization] can be derived.” (ibid., p. 82) Here, pragmatic (or performative) presuppositions function as ways of eliciting rational consensus, whilst propositional content says something about the truth conditions for a speech act. Habermas argues that in order to grasp the full functionality and context of meaning, propositional content is not enough, because it only looks at what “language says” (Finlayson, 2005, p. 32). Pragmatic content is needed as well if we want to grasp the full meaning of linguistic acts, because it tells us something about the way language is used, or what “language does” (ibid.). Reaching a reasonable consensus is possible and desirable, Habermas believes, because “reaching understanding inhabits human speech as its telos” (Habermas, 1984, p. 287).

The moral principle of universalization also has an ethical variant – the principle of discourse ethics (D), which says that “[o]nly those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse.” (Habermas, 1990, p. 93) (U) must be seen as "a rule of argumentation and [...] part of the logic of practical discourses", whereas (D) "stipulates the basic idea of a moral theory but does not form part of a logic of argumentation." (ibid.) The discourse principle is practical and “already presupposes that we can justify our choice for a norm” – for which (U) is needed (ibid., p. 66). (D) tells us that the subject of consensus must be moral norms. Nevertheless, this does not tell us how a form of agreement can be reached in a moral discourse. This is where (U) comes in, since (U) tells us that agreement on norms is possible if those norms equally respect the interests and values of the participants in the discourse. In other words, (U) explicates the demand for universalizability, which is the standard by which we can determine to what extent a moral norm is valid or not (Rummens, 2007b, pp. 85-86).

2.1.1 The rules of discourse

In order to defend the universal validity of discourse ethics, Habermas uses a transcendental-pragmatic argument to reconstruct the pragmatic presuppositions that every participant in a practical discourse must necessarily accept (Habermas, 1990, pp. 82-86). He argues that these

---

4 For a detailed discussion of this justification, see Habermas, 1990, pp. 78-82.
presuppositions are unavoidable for people's speech acts and actions. Thus, a transcendental-pragmatic reconstruction formulates what pragmatic presuppositions people *necessarily* have to accept if they wish to participate in a discourse. Without these presuppositions, a cooperative attempt to reach an agreement on the basis of arguments is impossible, because a denial of these presuppositions would result in a *performative contradiction*; that is, a denial of those presuppositions would contradict the practice of the discourse itself (I will return to this below). Habermas tries to reconstruct the rational bases of people's actions and intersubjective communication. According to him, the moral principle (U) can be reconstructed into a number of discourse rules that comprise what Habermas initially called an "ideal speech situation" (Habermas, 1973, 1984). These discourse rules, initially formulated by Alexy (1990), are as follows (Habermas, 1990, pp. 87-89):

**On the logical level:**
1. No speaker may contradict himself
2. Every speaker who applies predicate $F$ to object $A$ must be prepared to apply $F$ to all other objects resembling $A$ in all relevant aspects
3. Different speakers may not use the same expression with different meaning

**On the procedural level:**
1. Every speaker may assert only what he really believes
2. A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so

**On the process level:**
1. Every subject with the competence to speak and act is allowed to take part in a discourse
2. a. Everyone is allowed to question any assertion whatever
   b. Everyone is allowed to introduce any assertion whatever into the discourse
   c. Everyone is allowed to express his attitudes, desires, and needs
3. No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in (3.1) and (3.2)

As can be seen, Habermas formulates the discourse rules on three different levels. He does this to distinguish separate levels of presuppositions of argumentations, based on writings by Aristotle (ibid.). The first level corresponds to the level of *products*: one function of argumentation is to produce arguments that are intrinsically cogent. With these arguments, we can accept or refute claims to validity. Therefore, the logical level is based on matters of cogency and does not contain any form of ethical presuppositions. The second level relates to *procedures*. On this level, arguments are seen as processes for reaching an understanding in order to test problematic validity claims. The rules on this level do have an ethical content, because they are
seen as common to actions or discourses aimed at reaching an understanding *an sich*. On the third level, argumentation is seen as "a process of communication that [...] must satisfy improbable conditions." (Habermas, 1990, p. 88, my emphasis). The rules on this level are defining features of the ideal speech situation and unique to discourse as opposed to other forms of argumentation. Every speaker should presuppose the conditions these rules state if he/she thinks he/she is engaged in an argumentation. According to Habermas, the conditions that are needed for such an ideal speech situation to occur can be defined by using a "systematic analysis of performative contradictions" (ibid.). A performative contradiction arises when the content of a certain statement contradicts the very presuppositions that asserting the statement entails. A simple example is the assertion 'I am dead': because a person must be alive to be able to say anything, he cannot be dead and make the assertion 'I am dead' at the same time. Habermas believes that if performative contradictions are avoided, only the quality of the argument and the motive of cooperatively seeking truth count in a discourse. He contends that if one denies any of the pragmatic presuppositions he formulates, this contradicts the argumentative practice itself. For example, a reasonable agreement cannot be reached when people use the same word but with different meanings (1.3), for this would imply that one person would mean something else than another person when "reaching an agreement". If the agreement means something different for different people, then there is no real agreement, because people do not know or understand what they are actually agreeing to in other people's eyes. Another example concerns sincerity (2.1): people must only say what they actually believe, because if they do not, other people will not be able to reach an agreement with them that concerns universalizable norms. If I do not say what I really believe, then how can it be possible that the norms that flow out of the agreement respect my beliefs as strongly as others' beliefs? An example on the process level is the absence of coercion (3.3). An agreement cannot be reached if people do not fully "agree" to the norms that are being developed. It may be that a person is coerced to agree to something that that person would also have agreed to if she were not coerced. However, the point is that this person loses the ability to freely decide for her own what she does agree to and what not. It is this loss of freedom that forms the problem. Agreement is seen as something somebody chooses to do, not simply something somebody does.

It is important that Habermas himself believes that these discourse rules are not *constitutive* for a discourse, but merely say something about its *form*. We cannot abide to all rules, because that would most probably not even be possible. For instance, demanding that a person only talk in a way that is not self-contradictory, coerced or lied seems very ambitious. People cannot always hold fully consistent belief systems themselves; sometimes, their values simply contradict each other. Furthermore, the total absence of coercion seems to be utopian, because power always (partially) defines the relationships people have and because how people
behave in groups is quite different from how they behave when they are alone (cf. e.g. Janis, 1982; Sanders, 1997; Morriss, 2002; Lukes, 2005). Moreover, it is arguably so that people lie more often than they think. Sometimes people lie even whilst they themselves think they are telling the truth, for instance because they live in a state of false consciousness (cf. Geuss, 1981) or because they have a mental handicap. We should try to approximate these rules. The rules are not necessary prerequisites for a discourse, but, rather, tell us in what ways an actual discourse can take the form of an ideally organized discourse (Habermas, 1990, pp. 91-92).

What is also noteworthy is that the ideal speech situation attempts to come to a rational consensus through procedure; it says nothing about content, Habermas claims. A discourse must be seen as a group of people who intend to reach a rationally motivated consensus on certain moral norms that will be seen as universally valid. It is assumed that all people's interests are judged impartially and that, thus, only claims that are acceptable to all concerned are considered valid. Habermas calls this a "weak idea of normative justification" (ibid., pp. 92, 103-104).

It is important to note that Habermas develops a dialogical moral theory instead of a monological one (like Kant's or Rawls's). Individual reflection on whether someone assents to a norm is not enough for Habermas, because solving problems requires a cooperative effort, a consensus. "Only an intersubjective process of reaching understanding can produce an agreement that is reflexive in nature; only it can give participants the knowledge that they have collectively become convinced of something." (ibid., p. 67). The need for a real discourse in order to justify norms is one of the two basic assumptions of discourse ethics, the second of which is that "normative claims to validity have cognitive meaning and can be treated like claims of truth." (ibid., p. 68) Habermas deems this second assumption necessary because it is important to separate the validity of norms from the recognition of norms: norms can be recognized, but invalid, or vice versa. Habermas points out that practical discourse – and thus discourse ethics – is not about the real validity of norms, but what people expect to be justifiable, i.e. the norms that are recognized by the people, whether they can in fact be seen as valid or not (ibid., pp. 59-62). Moral norms are not some kind of descriptive assertions about an objective world of independent moral facts; their validity is constituted by agreement in an ideal, reasonable form of practical deliberation. This constitutes the cognitivistic part of Habermas's discourse ethics (cf. Rummens, 2007b, pp. 87-88).

According to Habermas, the dialogical character of discourse ethics is what makes his theory epistemologically superior to Rawls's, for instance. Rawls, Kant and many other authors have designed theories that require people to develop and justify moral norms through a monological process. In Rawls's case, they have to conduct a thought experiment and imagine what moral norms they would want to have if their particularities did not play a role in developing the moral norms. In Habermas's view, this is the wrong way to develop moral norms,
because imagining what people think or hold dear can never be as reliable as organizing an actual discourse in which people tell each other what they hold dear. We cannot possibly know or come to know everything about others’ preferences merely by imagining what they think, because others simply think differently. Moreover, people are situated in a certain particular context, with particular difficulties and preferences. For instance, it is very difficult, if not impossible, to imagine what a working-class family living in extreme poverty would hold dear for an upper-class, wealthy family, and vice versa. Or, to take another example, it is often difficult enough for a man to imagine what a woman wants/feels, and vice versa. This form a point of critique on the perspective of the “generalized other” that Rawls uses in his theory. Habermas (1994) as well as Benhabib (1992), among others, argue that we should not (and cannot) generalize about others precisely because they are situated in particular contexts. Habermas’s theory allows for this, which, in his opinion, makes it epistemologically superior to Rawls’s. I will return to this difference between Rawls and Habermas more elaborately in the next paragraph.

Finally, it should be noted that discourse ethics is a form of critical theory. It “doesn’t merely give information about how it would be rational for agents to act if they had certain interests; it claims to inform them about what interests it is rational for them to have.” (Geuss, 1981, p. 58) Perhaps Habermas’s critical theory is not as pessimistic as, for instance, Horkheimer’s or Adorno’s (cf. Finlayson, 2005), but it is still focussed on theory as well as practice. Habermas sees critical theory as a better form of theory than contemporary scientific theories, because they have an instrumental aim and are seen as separate from the objects they describe. They only try to observe the world instead of really understand it (Geuss, 1981, pp. 88-92). That is, they cannot provide us with a good account of all the complex, intersubjective patterns that exist in human relations and human life. Moreover, it tries to depict modern society as it is, but also tries to show us how it could (and should) be enlightened and emancipated. Habermas’s critical theory is reflective and self-referential (Geuss, 1981, pp. 55-59): the theory is part of the object-domain it describes. It must tell something about both its content of origin and content of application. The aims of Habermas’s critical theory are enlightenment (people are freed from false consciousness) and emancipation (people are freed from self-imposed coercion). Enlightenment and emancipation are possible through an increased capability of people to be self-reflective and to ask themselves whether their situation is still acceptable to them (ibid., pp. 58-59).
2.2 Habermas’s critics

2.2.1 Consensus and procedure

A first strand of critique on Habermas’s notion of discourse ethics focusses on the procedural character of discourse ethics. A part of this critique concerns the consensus (or agreement) that is the aim of discourse ethics. Although he was more positive towards the possibility of consensus in earlier works, Habermas emphasized in more recent works that rational consensus is impossible – both empirically and conceptually (Habermas, 1996b, p. 1518; Habermas, 1998a, pp. 365-369). Such a consensus is paradoxical, he argues, since when a consensus is achieved, discourse will stop: because of the working of discourse, it stops itself, because no discourse is needed anymore if there is a consensus. Thus, such a final consensus in fact contradicts the whole notion of discourse ethics (cf. Rummens, 2012). Thomassen (2010, p. 101) even argues that one might say that this view on a consensus in discourse ethics is a performative contradiction: “what makes communication possible (namely the implicit telos of a rational consensus) also makes it impossible and superfluous.” Thus, consensus could not be the goal of discourse, or at least not the only one.

Some critics connect another sub-problem to the problematic notion of consensus: the likeliness of reaching an agreement seems to decrease the more constraint-free or open a debate becomes (cf. Walzer, 1989; Chambers, 1993, pp. 231-233). This would imply that Habermas’s discourse rules do not foster, but rather impede the process of finding consensus. Certain persons in function (chairpersons, judges etc.) often forcefully end a debate, because it would escalate if they did not do so. People are often so convinced that their assertions after the debate are the same as the ones they had before. Moreover, some (temporary) differences of opinion will unlikely be resolved at all (cf. Rawls, 1987). Simone Chambers (1993) responded to this critique and defended Habermas’s discourse ethics by arguing that this kind of problematique only holds for single conversations; opinions change over time. This means that consensus is not something that occurs at one time and place, but is “the product of many single conversations” (Chambers, 1993, p. 232). Habermas acknowledges this in later work (1998a, pp. 365-369) and argues that we should not look at future agreement, but, rather, at how an ideal speech situation and a consensus would look like in the present. He also recognizes that these agreements can change over time. This seems to be a solution to the previous criticism: if we use consensus merely as a kind of temporary goal and acknowledge that consensus can be “broken” at any time, we have an incentive to keep on deliberating5.

---

5 See, for instance, Mouffe, 2000; 2005 and Rummens, 2012 for a more extensive approach regarding the possibility of “multiple concensi”.
Another point of critique concerning consensus is sometimes raised. Lyotard (1984) was one of the first prominent authors who have argued that Habermas's conception of discourse ethics implies an imposition of consensus: particularity or otherness of people who do not belong to the “mainstream” are ignored, since having consensus as the highest – or even only – goal of discourse would impose a consensus exactly where there is none. The latter – the absence of consensus – is exactly the reason why a discourse would be organized in the first place: to reach a consensus. However, if this consensus is reached, there is no real consensus, because it is a forced, an imposed consensus. Chantal Mouffe also recognized this problem and later added that the goal of Habermas's discourse ethics is wrong: reaching a rational consensus is impossible, but also unnecessary and even undesirable, because difference and dissent are the very things that characterize democracy (Mouffe, 2000, p. 48; Thomassen, 2007, pp. 27-33).

Arash Abizadeh (2005) argues that this is not necessarily the case, though. He argues that collective identity does not necessarily presuppose a tangible “Other”. The definition of a collective identity may also be based on a historical or hypothetical Other, he believes. Therefore, consensus can be reached in a discourse now, as long as the Other is defined in a hypothetical or historical manner. This provides Habermas with an answer to Mouffe’s point of critique.

2.2.2 Problems with the principle of universalization (U)

Some objections to Habermas's discourse ethics are aimed at the principle of universalization (U) specifically. A first problem is that only very few norms will be accepted as valid if (U) has such a wide scope of consensus (Finlayson, 2005, pp. 80-81, 87-88). If all people concerned must reach an agreement, the procedure seems very demanding. Only very general norms will be accepted, and only very few. First, Habermas responded to this point of critique by simply denying it. In later works, he conceives of it as a strength of his theory instead of a weakness. The general norms that are accepted as valid are, in his view, the most important ones. However, there is still a problem in Habermas's theory: why would discourse be so central to and important for safeguarding social order and social integration if there are only very few valid norms to solve conflicts through moral discourse? I believe this is a weak spot in Habermas's theory that has not yet been solved.

A second problem concerns the fact that Habermas's theory is dialogical (Finlayson, 2005, pp. 88-89). The problem with this is that a discourse may only include a small number of actual participants – regardless of whom the norm affects in total. Habermas does not provide us with evidence that a small number of people in a discourse are more likely to be correct in

---

6 This touches upon a prominent problem in political theory, viz. the “boundary problem” concerning the so-called “all-affected principle”. For extensive accounts of (and solutions to) this problem, see Goodin, 2007; Miller, 2009; Näsström, 2011; Verschoor, 2012.
validating/refuting a norm than only one individual. I believe that Habermas could argue that there are certain *degrees* of the benefits of a dialogical approach to morality. In this case, a discourse would become better the more people affected by the discourse would actually take part in the discourse. However, I will not fully develop this potential defence in favour of Habermas here.

A third problem is that Habermas's derivation of (U) is not fully non-moral. Some critics believe that the discourse rules he uses are not morally impartial; they are moral assumptions in Habermas's theory, although he claims principle (U) is justified on premises that are non-moral (Finlayson, 2005, pp. 89-90). For instance, rule 3.2c says that everybody must be allowed to express his/her attitudes, desires and needs. However, we have already seen that it is often so that not all people who are affected by a norm can actually take part in a discourse. Thus, in fact, rule 3.2c cannot be a rule of general discourse. Thus, the discourse rule is not performatively non-contradictory anymore (see Rummens, 2012 for a discussion of this in a slightly different way).

2.2.3 Impartiality and generalized versus concrete others

A different strand of critique on Habermas's discourse ethics focuses on the impartiality of the procedure of discourse ethics. Recall that, for Habermas, discourse ethics is only concerned with the way in which a discourse is organized, not with the content of the claims voiced in such a discourse. This means that people's ethical convictions are subjected to impartial moral norms, and that Habermas only focusses on the *procedure* of the creation of those norms, not the content of those norms. This is different from, for instance, John Rawls's theory of justice (Rawls, 1999a). Rawls develops substantive principles of justice by introducing a set of basic liberties. In his reply to Habermas, Rawls explains why, in his view, Habermas's theory is too metaphysical: he argues that a theory of justice ought only to concern *political* justice and institutions and not be reliant on a person's conception of the good or a philosophy of language (as Habermas's does, Rawls believes), because this would only lead to more disagreement (Rawls, 1995, pp. 133-138; Thomassen, 2010, p. 109). Habermas, however, responded to this by arguing that this is unavoidable and also undamaging if it is used in the right way (Habermas, 1998b, chapter three). Habermas asserts that, due to the many cultures there are today, substantive elements cannot be used in a moral theory; only a procedural conception of justice can help us, through determining what is just and what is not without touching upon the substantive preferences of the cultural plurality. Because of their different fundamental assumptions, Rawls and Habermas seem to argue on different levels.

But that is not the only point of quarrel between Rawls and Habermas. Although they both favour an impartial procedure for choosing principles of justice, they have a different
notion of impartiality as such. Rawls has a monological conception of impartiality: individuals use the original position as a thought experiment in their own minds, not in practical dialogue with others. By using this original position and by going back and forth, reflecting on norms and principles, a reflective equilibrium might be realized – at least it is striven after. For Habermas, impartiality originates in a practical discourse between individuals, in which they act as if they were in an ideal speech situation. It is dialogical, not monological. He responded to Rawls by rejecting the notion of an original position, because for Habermas, valid norms are only those norms that can be accepted by all affected individuals in an actual discourse (Habermas, 1990, p. 67; O’Neill, 1997, p. 107). There is a “common will” idea connected to this: shared assertions are more than just the aggregated principles that all individuals have created, isolated from each other, in a monological way. Moreover, Habermas argues, an actual dialogue has the advantage of allowing for revision: people hear from others what they have to say about their own views and through a discourse, people can revise their convictions and personal principles (O’Neill, 1997, pp. 107-108). Differences between individuals are thus allowed, or even used to reach an agreement. Rawls’s original position does not account for this, according to Habermas, because everybody is stripped from their particularities – everybody should be “as same as possible”.

This relates to the epistemological superiority of Habermas’s theory over Rawls’s mentioned earlier. This focus on difference is also related to the previous point of critique: Habermas thinks a philosopher should not be concerned with the content of a discourse (i.e. substantive principles or norms), but only with its procedure. We should let the participants in a discourse decide for themselves what kind of moral norms they want to create. Furthermore, the veil of ignorance itself causes a motivational problem: if people believe that they are forced to follow certain principles of justice after the veil has been lifted, they will probably not be motivated to follow the principles that have been created behind the veil of ignorance. Habermas believes that thinking of human beings as “generalized others” is not enough; there must be “concrete others”, with whom we formulate the principles of justice for our society. People do not only have some kind of universal capacity to be autonomous, but also have a particular view on the world and live under particular circumstances (Habermas, 1994). Although this was Carol Gilligan’s (1982) critique on Habermas in the first instance, Seyla Benhabib has changed the conception of generalized and concrete others in favour of Habermas’s theory. Gilligan based her critique on Habermas on the gender bias that she believed was in his theory, because “feminine” values like compassion or interdependence do not have a prominent place in Habermas’s theory. She connected the notion of the generalized other to masculine values and norms, and that of the concrete other to feminine ones (also see Schweickart, 1987; Benhabib, 1992; Tronto, 1993). Benhabib has broadened both notions by arguing that the generalized other comprises universal rights and formal reciprocity, whilst the concrete other concerns particularity and so-called
complementary reciprocity (individuals complement each other because they all have particular, unique characteristics) (Benhabib, 1986a, pp. 339-342; 1986b). Benhabib suggests that discourse ethics must be concerned with both the generalized and the concrete other. Habermas has used these suggestions in later works, in which he argues that discourse ethics allows for both sorts of “others”, because “the moral perspective is supposed to consider each and every voice.” (Thomassen, 2010, p. 106; cf. Habermas, 1990, pp. 175-181; 1993, p. 153) In this sense, Habermas has used important criticism to defend and strengthen his theory, which means, at least for some scholars, that the problem has been resolved. Again, this relates back to the epistemological advantage Habermas’s theory has, in comparison with Rawls's.

A different response to the claim that Habermas’s theory would be gender-biased is given by Wright (2004). He shows that these criticisms are incorrect and that critics read Habermas the wrong way; they fail to see, he claims, that Habermas allows for feelings not only in discourses themselves, but also outside of them: the boundary between ethical and normative expectations, he argues, is not fixed in discourse ethics; it leaves people to decide for themselves, through the discourse(s), where the boundary must be drawn (Wright, 2004, pp. 58-63).

Jean Keller (2008) goes even further than Wright “by using Habermas’s discourse ethics to develop a moral phenomenology of friendship.” (Keller, 2008, p. 160) In short, she argues that applying discourse ethics to friendship shows us that discourse also requires forms of assertiveness, receptivity and reciprocity, instead of only the cold, masculine, rationality-driven, contest-like discourse that some feminists think is what Habermas advocated (cf. Schweickart, 1987). Rationality and reason do have a prominent place in Habermas’s theory, she argues, but in order to be receptive, reciprocal and assertive, one must also develop emotional and social capacities (Keller, 2008, pp. 177-178). Thus, she defends the view that discourse ethics also accounts for, and even has the requirement to stimulate, “feminine” values and behaviour, although she also acknowledges that Habermas should have been clearer about this.

For others, like Stella Gaon (1998), the morality-ethics combination remains problematic. Gaon argues that discourse ethics needs a separation of morality and ethics, but that this separation, in turn, leads to two contradictory elements: on the one hand, discourse ethics must be formalist in order to provide a universalist account of morality; on the other hand, a moral theory must have something substantively moral about it, because else it would be trivial. These two elements are contradictory because discourse ethics must be “both ethically empty and yet normatively full.” (Gaon, 1998, p. 689)

In this vein, other authors also argue that there are still problems with Habermas’s notion of impartiality. Communitarians like Charles Taylor (1994) and Alasdair MacIntyre (1985), for instance, have argued that, although Habermas claims that his discourse ethics is universally valid, it is just one amongst many perspectives on ethics and morality. Related to
this, Habermas is sometimes accused of having developed a Western-, liberal democratic-oriented theory (Rummens, 2007b, pp. 93-94; Thomassen, 2010, p. 108). Habermas does not seem to realize that his theory is, in fact, also just a particular theory with a particular conception of the good – which is unavoidable in these communitarians’ opinions. Liberals, Rawls amongst them, have taken this argument further, because they do not think that this particularity is unavoidable (Erman, 2007). They contend that moral norms concerning justice can be separated from ethical conceptions of the good. And this relates back to the first point of critique that Rawls has on Habermas’s discourse ethics.

Albrecht Wellmer (1991) has pointed out another problem in Habermas’s discourse ethics: in his view, “[u]niversalism seems to overtax the limited capacities of our rational faculty and to necessitate the operations of a divine intellect.” (Habermas, 1993, p. 35) Habermas initially responded to this by arguing that deciding what we ought to do – ethics – does not necessarily have to deal with what we ought to do now or in this particular place, under these particular circumstances. For Habermas, it is a general question for which the solution is given through the principle of universalization (U) (ibid.). A stronger possibility for Habermas to defend himself has been given by Klaus Günther (1993), who suggested that Habermas should distinguish between two sorts of discourses: a discourse of justification and a discourse of application. The first discourse would be organized to justify the moral norms under consideration and reach a general consensus. The question “Is it just?” (i.e. “Is it equally good for all?”) is central to this discourse. The second discourse would be held after the first one and would concern the question “Is it appropriate?” for concrete, particular cases. Thus, the discourse of justification is about coming to agreement on the general justifiability of moral norms, whilst the discourse of application concerns applying those norms to particular circumstances and investigating whether they are appropriate. Günther proposes this separation of two discourses because he asserts that norms should also be applicable, not only justified: a generally justified moral norm that cannot be applied to reality has no use. Moreover, using two different discourses takes the pressure of identifying all possible implications of a norm off the justification discourse. Furthermore, the norms that have been justified in the justification discourse are not final anymore: the discourse of application allows for norms to be rejected because of their practical inapplicability, even if they are generally justifiable (Thomassen, 2010, p. 103). Using these two discourses does not only solve the problem that Wellmer pointed out, but also allows for a stronger duality of both the generalized and concrete others in discourse ethics. This is because the discourse of justification concerns a realm that holds for everybody, namely purely the justification of morals norms, whereas the discourse of application is applied to a particular temporal and spatial context. Because it is situated in this particular sense, the concrete situation of the people taking part in the discourse of application is
better accounted for than in a discourse (of application) that holds for all people, regardless of context. Habermas acknowledged that Günther’s suggestion makes sense and therefore incorporated Günther’s idea in his discourse ethics (Habermas, 1993, pp. 13, 35-39; 1996a, pp. 217-219).

This, however, creates a new problem, Bernstein (1995, pp. 222-228) argues. Bernstein asserts that when using two separate kinds of discourse, the discourse of application is only necessary because the discourse of justification is insufficient to account for the justification and the practical reflection of norms. The application discourse is thus used as a check for the norms that are justified by the justification discourse. This means that norms that have already been justified will never be absolute, because they can always be re-interpreted in the discourse of application. This seems to take away the aim of the justification discourse; we could also simply judge and potentially justify every norm in the concrete situations to which they have to be applied. Of course, this would mean that we do not need a discourse of justification and, in turn, that the discourse of application would be constitutive of norms. If this is so, then particularities will dominate the creation of norms, which makes Habermas’s theory culturally partial. This leads us back to the problem of generalized versus concrete others, or cultural pluralism problems. According to Habermas, it is right to account for particularities and context in the application of the norms, but not in their justification, because the norms must be universally valid. Therefore, paying attention to particular circumstances must be restricted to the discourse of application.

Especially if one looks at the relatively recent work of Rienstra and Hook, it seems as if the problems Wellmer discussed earlier are still present. Rienstra and Hook (2006) argue that discourse ethics asks too much of agents. In their article, they show that empirical and psychological studies suggest that the sort of rationality Habermas wants the participants in a discourse to have is not feasible, because the agents he talks about are simply nowhere to be found in current societies. Moreover, they argue, precisely those institutions that could make such a rational agent possible would contradict Habermas’s theory: “[t]o attain the institutional requirements of successful deliberation one must break from Habermas's procedural preconditions concerning power and communicative symmetry. To ensure deliberative success, non-neutral agendas are required, and expert communication is demanded.” (Rienstra & Hook, 2006, p. 336) Although Habermas could respond to this by saying that two discourses can be organized, Rienstra and Hook’s argument poses a relevant problem to Habermas’s theory: what if people simply cannot be as rational or as prudent as Habermas suggests they are, even if two (or more) separate discourses take place? Is justifying and/or applying moral norms not too difficult for ordinary human beings? What if discourse ethics really “necessitate[s] the operations of a divine intellect”? (Habermas, 1993, p. 35)
2.3 The desirability of a deliberative democratic theory

The aim of this thesis is to devise a way in which the feasibility of Habermasian discourse ethics can be increased. Before I set out to create such a model, it should be discussed why we would want a deliberative democratic theory in the first place. I will formulate a short answer on this, first, by referring to the advantages of Habermas’s discourse ethics that were touched upon in the previous two paragraphs; and secondly, by discussing several views on the desirability of deliberative democratic theories from an empirical perspective.

First, the deliberative model, with its emphasis on human intersubjectivity, has an advantage over empiricist models, which think of human beings as acting rationally in an instrumental sense. According to Habermas, it is possible to define common goals and interests by participating in a reasonable discussion. Rummens (2007a, p. 21) argues that, empirically speaking, the empiricists’ assumption that people only act out of particular self-interest is untenable (also see Ackerman, 1989; Mackie, 1998; Druckman, 2004). Moreover, discourse ethics overcomes several conceptual problems, as Rummens points out. For instance, theories that assume instrumental rationality cannot explain human cooperation, since those theories would have to expect that the best choice people can make is to become a free-rider and thus benefit on others’ expenses. This cannot account for cooperative behaviour, so all forms of behaviour that cannot be seen as instrumentally rational are seen as irrational behaviour, which seems to be far from the truth.

The problems with instrumentally rational models can be overcome by thinking of political theory in terms of practical reasonableness, which goes beyond the narrow conception of instrumental rationality. In this view, it is believed that a reasonable discussion about the common good of a society can be organized (cf. Rawls, 1999a). Habermas believes this is possible without choosing to endorse either a liberal or a republican conception of autonomy and rights. Because private and public autonomy are equiprimordial (or co-original) in Habermas’s view (Habermas, 1996a, pp. 88-94, 121-123, 134-135, 161-165; 1998b, pp. 298-303; Rummens, 2007a, pp. 25-33; Baynes, 2009, pp. 542-546), we do not have to choose one and

---

7 Specifically, this says that a person has private autonomy because she can decide for herself how she wishes to lead her life, but a person also has public autonomy because she can decide, in agreement with other citizens, what laws their community should have, and what the common good is. The rights we have as private persons are designed by ourselves, together with the rest of our community, and the public autonomy we have follows from our rights as private citizens, because people must be able to decide for themselves what laws should be developed for them as private persons. The possibility to take part in the democratic decision-making process on an equal footing, in turn, presupposes the citizens’ private autonomy. This way, neither private nor public autonomy is more constitutive than the other, Habermas asserts. Additionally, the tension between universalism and particularism is removed. The recognition of citizens’ public and private autonomy is a universal demand that a democratic community must guarantee. Conversely, only the citizens who take part in the decision-making procedure can determine what that demand of recognition means exactly, because citizens are situated in a certain temporal and spatial context. The meaning of the demand for recognition will be heavily influenced by this context. Also see Rummens, 2007a, p. 31.
abandon the other. Moreover, we are able to account for universality as well as particularity; not only through the co-originality of democracy and rights, but also through the attention that is paid to the “concrete other” in Habermasian discourse ethics. After Habermas, following Günther (1993), had introduced two kinds of discourse in his theory instead of one (a discourse of justification and a discourse of application), this attention to the concrete other became even more extensive: the more general discourse of justification safeguards universalism, whilst the discourse(s) of application is/are sensitive to the context in which particular people are situated.

Aside from this, Habermas’s theory allows us to grasp the full functionality and context of “meaning”, namely by looking beyond propositional content (i.e. what language says). Habermas's emphasis on both pragmatic content (i.e. what language does) and propositional content gives us this opportunity. This emphasis on practice is important for Habermas, because discourse ethics is a critical theory (Geuss, 1981). It tries to understand and enhance the world, not just say something about it from a dissociated, object-independent standpoint. By being a part of the subject one studies oneself, one can really grasp what people think and account for the particular context they are situated in. An advantage related to this is the dialogical character of discourse ethics. Because it allows for a more practical and direct approach by organizing an actual discourse to design and justify moral norms, discourse ethics gives people a better understanding of their fellow citizens’ preferences and values. Because of this, it is argued, Habermas’s theory is epistemologically superior to monological theories like John Rawls’s.

Empirical studies also suggest that deliberative democratic models produce certain benefits. Chambers (1998), for instance, argues that, compared to contractarian theories (in her case Locke’s), deliberative democratic theories can account for diversity in a better way. This is because the latter are mostly open-ended and do not have to develop institutional rules or constitutional rights once and for all (like traditional contractarian theories do). Delli Carpi and colleagues (2004), in their review of the empirical literature on deliberative politics, acknowledge that deliberation, as well as other forms of discursive politics, can be individually or collectively beneficial, but that its impact is highly dependent on the context. Jason Barabas (2004) finds that Chambers’s emphasis on the openness of deliberative models is justified, because deliberation increases the level of the participants’ knowledge and alters or softens strongly held opinions. However, his findings also suggest that these effects of deliberation strongly depend on the willingness of deliberation participants to be open to others’ opinions as well as on the quality and the diversity of the assertions in a deliberative setting. Finally, Persson et al. (2013) conclude from their empirical study that deliberation seems to promote the perceived legitimacy of democratic procedures. Still, this result does not always hold, for the effect of direct referendum voting seems to overshadow the effect of deliberation on perceived
legitimacy. Moreover, the effect of deliberation on perceived legitimacy strongly depends on people's perceived influence on decision-making procedures.

However, there are also authors who do not believe that deliberative democratic approaches – at least in the form they have been presented so far – are fruitful. Sanders (1997), for example, argues that deliberative democratic theories do not weaken, but strengthen the hierarchical power relations that are present in contemporary society. She does this by looking at citizens' juries in the US. For her, it seems that somebody's success in deliberation depends, at least partially, on his/her race, gender, role in the jury etc., which naturally conflicts with the rules set out for an "ideal discourse". Mutz (2008) argues that deliberative theories are unfalsifiable, because they use too broadly defined concepts and too many necessary and/or sufficient causes for the outcomes they "predict". Therefore, she argues, we cannot consider deliberative democratic theories as "grand theories". Mutz believes we should disentangle deliberative theory into a number of "middle-range theories" in order to be able to capture and measure the products and stimuli of deliberative democratic dialogue. Thompson (2008) believes that there is a future for deliberative democracy in both the empirical and the normative realm, but only if the elements of deliberative democratic theories are more clearly defined; the conflicts between those elements are investigated; and the structural relationship in deliberative systems is paid more attention to.

As can be seen, not all authors contend that deliberative democratic models offer us a fruitful and correct way of thinking about morality. Still, many authors do see the merits of these models, both empirically and normatively speaking. These latter authors have shown us that deliberative theories have several benefits compared to other theories, some of which resolve problems that are often seen as important for political theories – private vs. public autonomy and universalism vs. particularism, for instance. It seems that trying to make Habermasian discourse ethics more feasible is worthwhile, for this would bring us another step further towards overcoming prominent problems in political theory.
3 A BIPARTITE CONCEPTION OF FEASIBILITY

Thus far, I have focussed on Habermas’s moral theory of discourse ethics and several points of critique on it. Now, I will focus on the notion of feasibility. However, I have not yet argued why feasibility should at least be taken into account when formulating political theories. I will first elaborate on the need for a feasible discourse ethics. After this, I will explicate how I will use the notion of feasibility in the next part of this thesis. What do we mean when we say a theory is feasible or not? What criteria does a feasible theory have to satisfy? In other words, what is feasibility? I will mainly use an illuminating article by Gilabert and Lawford-Smith (2012) to answer these questions, because it provides us with a clear and many-sided discussion of the elements of feasibility. I will describe, similar to what is done in the article, what concepts are often thought of as being important to feasibility. Finally, I will present the account of feasibility that has been developed in the above-mentioned article, altering it in some minor ways.

3.1 The need for a feasible discourse ethics

One could ask why political theorists should care about feasibility in the first place. As Gilabert and Lawford-Smith (2012, p. 810) point out, John Rawls’s idea of a “realistic utopia” can be quite useful here (Rawls, 1999b; 2001). Following Rousseau (cf. Rousseau, 2003, book 1, p. 1), Rawls (1999b, pp. 12-13) thinks that for a liberal conception of justice to be realistic, it must rely on the actual laws of nature and achieve the kind of stability those laws allow, that is, stability for the right reasons. It takes people as they are (by the laws of nature) and constitutional and civil laws as they might be, that is, as they would be in a reasonably just and well-ordered democratic society.

Moreover, the first principles of such a conception must be “workable and applicable to ongoing political and social arrangements.” (ibid., p. 13)

Rawls believes that political theories must contain elements of this kind of realism as well as utopian/idealistic elements⁸. On the one hand, Rawls asserts, a political theory must prescribe desirable or morally appealing social arrangements to us. Of course, suchlike arrangements can be “utopian” if they are not strongly similar to the social arrangements that exist at the moment. Rawls contends that this is not a wrong, but, in fact, often a good thing: exactly because the social arrangements the theory aims to achieve are so far removed from the

---

⁸ Also see Carens, 1996 for an elaboration on the trade-off between ‘ought’ and ‘is’, i.e. realism and idealism.
ones that actually exist, we are able to “look into the mirror” and critically assess the current state of affairs. On the other hand, this critical potential is not unlimited: we must also take into account the extent to which we “realistically” expect to actually achieve the outcomes that the normative theory demands. This touches upon the feasibility of a theory. Rawls believes that “political philosophy is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in doing so, reconciles us to our political and social condition.” (Rawls, 1999b, p. 11) Rawls sees the capacity for a theory to be realistically utopian as one of the roles that political philosophy can have as a part of the public political culture of a society. Nevertheless, he believes it is difficult to fully grasp the notion of feasibility or the conditions of our social world. “[T]he problem”, he contends, “is that the limits of the possible are not given by the actual, for we can to a greater or lesser extent change political and social institutions, and much else.” (Rawls, 2001, p. 5) But still, Rawls believes, as do I, that the “realistic” element of a realistic utopia needs to be considered, because it has such a central role in the dialogical process between “ideal” and “non-ideal” theory described above (also see Mutz, 2008, pp. 523-524). By considering the feasibility of theories, we can judge whether the source of discrepancies between a theory and reality lies in the former, the latter or both.

There is another reason why feasibility, in Habermas’s case specifically, should be important. As we already saw at the beginning of the present chapter, Habermas’s moral theory of discourse is a critical theory (Geuss, 1981; Finlayson, 2005). This means that the theory does not only say something about objects in a world the theory is not a part of. A critical theory is part of the object-domain it studies and not only observes a state of affairs, but tries to understand it. Moreover, Habermas’s theory not only tries to report what modern society “looks like”, but also provides a way in which it could be improved. Critical theory consists of three main parts (Geuss, 1981, pp. 76-78): first, a part that shows us that the proposed societal transition as well as the final state that is aimed at – emancipation and enlightenment – are theoretically possible. This part is to guard against utopianism. Secondly, there is a part that shows that the transition towards the proposed final state is practically necessary. This part explicates why the transition will lead to more emancipation and enlightenment, and is thus practically necessary. It also avoids scientism (the assertion that all knowledge one possesses

9 The other roles of political philosophy Rawls enumerates are the following: 1) the practical role, “to focus on deeply disputed questions and to see whether, despite appearances, some underlying basis of philosophical and moral agreement can be uncovered.”; 2) the role of orientation, to “contribute to how a people think [sic] of their political and social institutions as a whole, and their basic aims and purposes as individuals; and 3) the role of reconciliation, to “try to calm our frustration and rage against our society and its history by showing us the way in which its institutions [...] are rational” (Rawls, 2001, pp. 1-3).

10 See Mutz, 2008 for an attempt to bring political theory and empirical science together; see Wolff, 2011 for an attempt to apply political theory to specific cases in reality instead of merely thinking about it in a utopian, “ideal” sense; see Valentini, 2012 for an extensive conceptual elaboration on the notions of ideal and non-ideal theory.
about society is of a scientific nature). A third part states that the transition towards the final state can only be achieved if people use the critical theory as their source of “self-consciousness” and act on it – in other words, if people adopt the critical theory and also let their actions be guided by it. If all three parts are present, the critical theory can be “confirmed”/corroborated.

The first part is not exclusively found in critical theories, for it can also be a part of a scientific theory. Exactly this part says that critical theories should be “empirically accurate” (ibid., p. 79). For Habermas's discourse ethics, this not only means that the current state of society as well as its final state should be empirically corroborated and clear, but also that the means that are used to get to the final state – i.e. discourse and discourse rules – should be feasible. For else, the final state would not even be attainable and the critical theory would not be adopted by people who are in a state of freedom/absence of coercion.

Feasibility becomes even more important if one considers Habermas’s conception of universality. For Habermas, ideal role-taking is not done from a third-person perspective (Finlayson, 2005, pp. 84-85). Precisely because the “universalizer” him-/herself is part of the object-domain he/she is investigating, he/she must adopt a first-person perspective on morality. People who take part in moral discourse are real, concrete people, in a real lifeworld, who are guided by the rules of discourse. Because every individual’s perspective is necessarily finite, an actual discourse must take place in order to attain the ideal of an “unlimited communication community” (Habermas, 1993, p. 51) and, consequently, to be able to justify norms (also see Habermas, 1990, p. 94). Moreover, agents in such a discourse must be free, uncoerced, sincere etc. in order to develop this kind of communication community from their individual, finite perspectives. Thus, in order to attain (or approximate) the ideal discourse situation, an actual discourse must be organized – and its discourse rules must, at least to a sufficient degree, be feasible. This importance of the organization of an actual discourse is also emphasized by the inherently dialogical character of a discourse: people must communicate with each other if a discourse is to take place (Finlayson, 2005, p. 85). Because Habermas’s theory strongly depends on the possibility and probability that an actual discourse can be organized, the theory becomes problematic, or even useless, if such a discourse cannot be organized. Therefore, it must at least be possible to organize an actual Habermasian discourse if we want to prevent Habermas’s theory from being discarded in the first instance.

Finally, a more general reason why a deliberative democratic theory should be feasible can be given. For a theory to be convincing, it must possess a sufficient degree of consistency. On the one hand, theoretical consistency is important, because if certain parts of the theory – assumptions, conclusions, procedures etc. – are not compatible or even conflict with one another, the theory is often regarded as unconvincing. On the other hand, consistency between
theory and practice is also important. The assumptions in a theory should be possible and reasonably probable to appear or exist in "the real world", as well as the processes and the outcomes the theory describes or predicts. A theory that is only theoretically and not practically consistent cannot make the switch from theory to practice (because of feasibility problems^{11}); and exactly making this switch is what people often try to do with a theory. They do not only want to know what it means in a theoretical sense, but also what it implies for their reality. Arguably, full consistency between praxis and theory cannot be attained, and so the degree of consistency between them is important: in this realistic view, we should try to make the gap between 'is' and 'ought' (or theory and practice) neither too small nor too large (Carens, 1996).

Some idealistic theorists disagree and think that critical theory does not have to be feasible or practicable, because by being unfeasible, it does exactly what it is meant to do: be critical of reality and show what reality should look like (ibid.). They believe that "[t]heory challenges political reality." (Thompson, 2008, p. 499) However, this attitude has become less dominant among theorists, because many now also recognize that empirical research and practical approaches can tell us something about the feasibility and quality of deliberative models (cf. Bohman, 1998; Dryzek, 2007). Be it, perhaps, as a mere check for the assumptions or phenomena a theory describes or predicts, a theory's feasibility must be taken into account. The question remains to what extent a theory must be practically consistent, that is, how large a gap between 'is' and 'ought', between realism and idealism, is acceptable. Following Carens (1996), I think that every political theory must have both realistic and idealistic elements. Nevertheless, I focus on feasibility considerations in this thesis, because Habermas often seems to neglect these considerations. That is, I assert that Habermas has incorporated too strong a form of idealism in his discourse ethics.

3.2 Important features of feasibility

As Gilabert and Lawford-Smith (2012, p. 809) state, "[v]ery roughly, some state of affairs is feasible if there is a way we can bring it about." But, naturally, this is too vague and general a definition. Let us look at what several authors have deemed important elements of feasibility.

A first element of feasibility concerns momentum. Inspired by Hawthorn (1991, p. 158; also see Tetlock et al., 2006), Gilabert and Lawford-Smith argue that some events – be they situated in history or future – cannot be assumed to have been otherwise, or become otherwise, than in reality, because they possess an inevitability that makes it unlikely, or even "fantastical" to assume such a thing. Conversely, some events will never be possible or have never been

^{11} This will be touched upon more in-depth in the remainder of the current chapter and in the next chapter.
possible, either physically or contextually speaking. Gilabert and Lawford-Smith (2012, p. 810) give two examples:

someone might argue that Karl Marx had such an impact upon people's thinking, and Russia's political leaders were so ideologically driven, that it was bound to be the case that Russia ignored Marx's criterion for the kind of country 'ready' to instantiate communist ideals, and went ahead with its disastrous attempt. To transpose the point about counterfactual histories on to counterfactual futures, note that we are in a world still reeling from the failures of certain kinds of political 'experiments', like communism in Soviet Russia, racial purity in Nazi Germany and racial assimilation in colonial Australia. A future in which these experiments have succeeded is 'fantastic' because we cannot expect people simply to forget political history.

Put simply, we ought not to assume inevitable or impossible things, either physically speaking or contextually speaking.

Touching upon a second element of feasibility, Mark Jensen (2009) argues that there are four jointly necessary and sufficient conditions for "practical possibility", namely logical consistency (or coherence); the non-violation of physical laws (or nomological consistency); the use of history as a fixed point in theorizing (we almost always begin theories with the here and now as a starting point, and then look at what could be changed to make the world more perfect); and the reflection on natural human abilities (Jensen, 2009, pp. 170-173; also see Gilabert & Lawford-Smith, 2012, p. 811). Concerning the first condition, Jensen (2009, p. 170) argues that an example of logical inconsistency would be "something like M.C. Escher drawing a square circle." A simple example of a violation of a physical law is that a certain theory expects all people to be able to jump 100 feet high without assistance. Although such an ability is formally consistent, it is simply impossible in terms of physical constraints. Concerning the third condition, Jensen (ibid., p. 171) formulates the following example:

While it is possible in some sense that Locke lived in the 10th century instead of the 17th, we do not take this into account in formulating our socio-political ideals. We regard the history of the world as a fixed point in our theorizing. In fact, in most cases it serves as our starting point: we begin with present historical circumstances and describe how they should be changed in order to achieve a more perfect world.

Jensen argues that although all four conditions are important for practical possibility, the first three are not often violated, whereas the fourth one is (at least more often). He subdivides natural human ability into three categories: *synchronic*, *direct diachronic* and *indirect diachronic* ability. To have a synchronic ability is to be able to perform a certain action at this very moment. If a certain theory demands that I clap my hands right now, then I will have a synchronic ability if I still have two hands and if my hands are not tied or constrained in another way. When
somebody has a direct diachronic ability to do something, this means that he/she can perform this action either now or at a later moment. If I have two hands that are unconstrained and will certainly be unconstrained for the next seven days, then I will have a direct diachronic ability to clap my hands either now or at another moment during the next seven days. Finally, an indirect diachronic ability is an ability that can be performed either now or later, but requires a certain previous action to enable the actor to perform it. We see, then, that a previous action always has to be performed before the indirect diachronic ability can be used. This last category of natural human ability shows us that feasibility is not always concerned with what can be done now. For instance, I have an indirect diachronic ability to clap my hands if I have both hands but my hands are tied or constrained in another way. In this example, my hands must be untied either by myself or other people before I will be able to clap them. Thus, clapping my hands presupposes an action through which my hands will be untied. Following the analogy, if I have only one hand, or even none, then I cannot clap my hands right now and will not be able to clap my hands later. Thus, I do not have any form of an ability to clap my hands.

Thirdly, authors have emphasized accessibility and stability as important elements of feasibility (Cohen, 2001; Barry & Valentini, 2009). This is because they believe that principles themselves cannot be feasible (or infeasible); only the ways in which they are implemented can be. Such an implementation option is accessible when there is a clear way through which it is possible to arrive at the preferred situation (that is, the situation whose feasibility is being judged). "There must be a way to bring the state of affairs about." (Gilabert & Lawford-Smith, 2012, p. 811) For instance, if a theory’s goal is that every person in the world has the same IQ level because of equality concerns, there does not seem to be a way in which this can be attained – at least not in the direct future. Secondly, a preferred situation must be stable (or at least be likely to be stable), because arriving at a situation that will change shortly after arrival does not make the preferred situation more feasible. If a preferred situation dictates that all the people in a certain country have the same amount of money, this situation is highly unstable: because of trade mechanisms, inflation and other factors, people will have different amounts of money after the preferred situation has been reached – assuming that such a preferred situation is possible in the first place. A proposed state of affairs is stable if it lasts for a "reasonable length of time" and it becomes more stable the longer this lasts (ibid., p. 813). Therefore, an outcome must be (likely to be) stable.

Finally, it is important to note that feasible states of affairs do not demand anything of people that surpasses the limits of human capacities, natural resources or human psychology (ibid., pp. 811-812). This touches upon the well-known demand of "ought implies can": if people

---

12 The use of artificial hands could enable me to clap my hands again. However, I will not discuss this further, because I think the analogy is clear enough to point out the differences between the forms of natural human abilities that are mentioned.
simply cannot bring something about that a theory prescribes, they subsequently cannot be obligated to bring it about, for they cannot. An example is that people cannot consider all consequences of their actions, because there are too many consequences; because they would lose their mind in thinking about the consequences; because after actually determining the consequences, the situation will have changed and the consequences will probably have changed as well etc.

3.3 An account of feasibility

Gilabert and Lawford-Smith (2012, p. 812) state that there are three main questions that must be answered to say something about the feasibility of a state of affairs: 1) “Feasible for whom?”, which concerns the actors that will judge of a state of affairs for its feasibility; 2) “Feasibility of what?”, which concerns the feasibility of (the path towards) a certain outcome; and 3) “Feasible when and where?”, or context in both a temporal and a spatial sense.

These authors also assert that feasibility considerations have two roles: first, they can rule out infeasible policies; secondly, they can compare different policy proposals (ibid., p. 812).

In my opinion, feasibility considerations have a third, more positive function as well, viz. to support political theories. In the first function, feasibility considerations are a kind of “control mechanism” for theories in general and philosophical and highly “utopian” theories specifically: if theories are simply not applicable, they are ruled out. In the second function, theories are compared to establish certain degrees of feasibility per theory. Thus, the second function is focussed on the probability of theories being feasible, in comparison with other theories. The third function of feasibility considerations is that they can also establish a degree of feasibility of a certain theory without comparing it to other theories. If a theory survives the “possibility tests” (function one), evidence can be gathered in favour of, or against, supporting the practicability of the theory. This function shows us that feasibility considerations cannot only rule out or compare theories, but also sustain them, even without comparing them to other ones.

Let me give an example of how these functions work. Suppose we investigate a theory that prescribes a form of socio-economic redistribution. If the theory prescribes attaining full redistribution, we could say that it does not survive the “possibility test”, because attaining full

---

13 In this paragraph, I only deal with creating an account of feasibility. I am not yet explicitly concerned with how to increase the level of a specific theory's feasibility – in this case Habermas's. I will touch upon this later.

14 There are two distinct elements of the answer to this question: on the one hand, there are the actions that can (or cannot) be performed in order to bring about a certain situation (an outcome); on the other hand, there are the outcomes that can (or cannot) follow from these actions. If one of these elements is feasible, this not always implies that the other is also feasible (for instance, there can be multiple pathways to the same outcome) (cf. Gilabert & Lawford-Smith, 2012, p. 812).
distributive equality is impossible (cf. Nozick, 1974; Walzer, 1983). People do not have the same abilities, talents or chances in life, and many forms of bad luck cannot be fully compensated – to mention but two of the difficulties with full distributive equality as a goal. However, if a theory prescribes that only a limited amount of resources should be distributed in a limited amount of cases (take, for instance, Rawls’s difference principle (Rawls, 1999a, pp. 7, 13-14, 96)), and if this form of distribution seems to be possible, then the probability of attaining such a level of redistribution becomes important. So far, Gilabert and Lawford-Smith would agree; but they would also suggest to compare this theory to other theories at this point, whilst I think that this must not necessarily be done to assess such a probability. For instance, assume that one of the elements of the theory under consideration is that it becomes more feasible to the extent that a country’s economic growth is higher. If we then look at a country that has a very low level of economic growth, or even none at all, we have a strong incentive to conclude that this lowers the feasibility of the theory in a general sense, but also of the theory in that country specifically. In contrast, if we look at a country that has a high level of economic growth, the theory seems to be more feasible in that country, as well as generally speaking. We have not compared the theory to other theories, but in spite of this, we have been able to make inferences concerning the degree of feasibility of the single theory we have under consideration. This is what characterizes the third function of feasibility considerations. Comparing a theory to other theories is prudent, but that does not mean that it should necessarily be done to assess a theory’s feasibility.

The authors’ account of accessibility is strongly connected to the roles they believe feasibility considerations have (Gilabert & Lawford-Smith, 2012, pp. 813-815). They explore the meaning of accessibility by looking at its antonym: inaccessibility. In other words, what forms of constraints determine if a path to an outcome is accessible or not? The authors distinguish between hard constraints and soft constraints. The former type of constraints always forms a constraint; it is a dichotomous, deterministic kind of constraint. Hard constraints do not permit a degree of feasibility: if hard constraints make a certain course inaccessible, that course is simply impossible. Thus, hard constraints can be used to rule out theories. Examples of hard constraints are logical, nomological and, arguably, biological and psychological constraints. Logical constraints concern logical consistency. Nomological constraints are constraints that violate physical laws. Assuming that one thousand normal-sized people fit into a room of twelve square feet that is of normal height, for example, is nomologically impossible and therefore forms a hard constraint on a theory that makes this assumption. Biological and psychological constraints are (the lack of) human capabilities and elements of “human nature” that make the feasibility of the (path to the) preferred outcome impossible. For instance, we cannot expect people to speak three (spoken) languages at the same time, since the human body cannot be used for such a thing.
In contrast, soft constraints have a probabilistic element: they are a matter of degree, and their limits are neither absolute nor permanent. They make certain outcomes or actions comparatively more or less feasible, which makes them very suitable for the second and third roles of feasibility considerations.15 Whereas hard constraints are relatively permanent, soft constraints often have a high level of changeability: they can vary, whether this follows from human actions or not. Soft constraints are "subject to dynamic variation: not everything that is less feasible now […] need be as infeasible later." (ibid., p. 814) Examples of soft constraints are economic, cultural and institutional constraints. Economic constraints could be inequality of resources or low employment rates, for instance. Examples of cultural constraints are language barriers, religious diversity or the separation of men and women in many social situations or spheres of life. Institutional constraints could be the lack of transparency of institutions or the lack of the potential of an institution to change over time. Gilabert and Lawford-Smith also point out that motivational constraints might be important soft constraints for feasibility, but do not use these constraints in their article. I assert that motivational constraints are important for feasibility considerations. If people are motivated to do something, for whatever reason that may be, they will often put more energy into it. We can see this in schools, for instance: if students are eager to learn about a subject they have a high level of affinity with, they will put more energy into actually learning about the subject. If I take pleasure in studying plants and animals, for example, I am motivated to study biology. Precisely because I like to know more about plants or animals, I will probably excel in this subject; for not everybody will have the same preferences or interests as I have. Because of this, studying biology is not only likely to be easier and more satisfying for me than for others, but I will also probably learn more about biology than others. In this simple example, this makes studying biology more feasible for me in general, and more feasible for me in comparison with others.

Nevertheless, motivational factors seem to relate more to the desirability of a theory than the other types of constraints. I admit that desirability can play an influential role in assessing the feasibility of a theory, but this is not the main theme of this thesis. I am concerned with investigating (and enhancing) the feasibility of Habermas’s discourse ethics in this thesis. However, as we have seen, Habermas’s critical theory has the aim of emancipating and enlightening people. People will probably, at first, be reluctant to change anything, because they

---

15 In this thesis, soft constraints can perhaps be best seen as what Bennett (2010, pp. 210-211) would call a straw-in-the-wind test. This kind of test affirms the relevance of a certain hypothesis or theory if it is passed and suggests the possible irrelevance of a hypothesis or theory if it is failed. However, it cannot definitively confirm/corrobore or refute a hypothesis or theory, because the test is not strong enough for this. It can merely strengthen or weaken expectations of a certain theory or hypothesis. By contrast, hard constraints can be regarded as hoop tests (cf. Bennett, 2010; Goertz & Mahoney, 2012). These tests require a certain piece of evidence to be present in order for a certain hypothesis to be valid. Hoop tests relate to necessary conditions: conditions that must be present for a certain outcome to occur.
have not been enlightened yet and live in a state of “false consciousness” (Geuss, 1981). If people’s motivational considerations are so strong that a transition from the current situation to a certain outcome would imply too much resistance to actually be carried out, this has to be taken seriously; especially in Habermas’s case. Therefore, we must take into account motivational constraints when determining to what extent a theory is feasible.

The authors assert that soft constraints play an important role in feasibility considerations. Of course, they are malleable and therefore seem to be less “powerful” than hard constraints, but this does not mean that we should simply ignore or dismiss soft constraints. They influence feasibility and can form serious obstacles for certain actions or outcomes. The idea is that we can combine hard constraints, which approach feasibility considerations in a binary sense, with soft constraints, which do so in a scalar sense. Consequently, we have two ways of testing the feasibility of theories in terms of accessibility. The two different tests relate to three stages that can be used to determine the feasibility of a theory, the authors believe (ibid., pp. 819-821). They argue that in judging the feasibility of a theory, we must first investigate whether certain hard constraints are present. If these are absent, we must look at the stability of the pathways and outcomes that are under discussion. Stability is interpreted in a scalar sense, in terms of soft constraints. The third step is to assess the accessibility of the actions and outcomes. This is also done by looking at soft constraints. So we start by looking at “the big picture”, i.e. ruling out theories that are simply impossible, which is followed by a more particular, situation-specific analysis of the accessibility and stability of a policy proposal.

On the basis of the feasibility considerations presented above, we can construct two schemes (see Schemes 1 and 2 below) to represent the steps that need to be taken when determining the feasibility of a theory. Furthermore, we can formulate two separate general definitions of feasibility\(^\text{16}\) – one for hard constraints (HC), one for soft constraints (SC) – that, taken together, constitute what is meant by the notion of “feasibility”:

\[
\text{HC: it is feasible for an actor to (perform an action to) bring about a state of affairs in a certain temporal and spatial context if this is not incompatible with any hard constraint}
\]

\[
\text{SC}\(^\text{17}\): it becomes more feasible for an actor to (perform an action to) bring about a state of affairs in a certain temporal and spatial context according as the probability of that outcome to occur, given soft constraints, increases}
\]

\(^{16}\) I have roughly based the way in which these definitions are formulated on the two tests for feasibility that Gilabert and Lawford-Smith (2012, p. 815) have developed.

\(^{17}\) It should be noted that, concerning soft constraints, I assume people try to (perform an action to) reach the preferred outcome. Also note that this assumption does not concern motivational factors, since these are concerned only with whether an actor wants to (perform and action to) reach the preferred outcome – inasmuch as this actually is the preferred outcome for an actor.
As we have seen, it is necessary to look at HC first. We can only look at degrees of feasibility (SC) if an (action resulting in the) outcome is free from hard constraints.

**Scheme 1: steps for determining feasibility – hard constraints**

| TYPE OF HARD CONSTRAINT                                                                 |  |  |
|----------------------------------------------------------------------------------------|  |  |
| Is the preferred outcome either inevitable or impossible in terms of momentum?         | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Does the (path to the) preferred outcome violate physical laws (Is it nomologically inconsistent)? | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Is the starting point from which the preferred outcome is to be realized situated at a historical place and/or time? | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Is the path that leads from the current situation to the preferred outcome logically inconsistent (or incoherent)? | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Is there a biological or psychological constraint on the (path to the) preferred outcome? | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Is/will the actor that is to bring the preferred outcome about (be) unable – in both a direct and indirect sense, in both the present and the future – to bring the outcome about? | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |
| Does the preferred outcome completely lack stability?                                   | YES | No further feasibility considerations will follow |
| NO                                                                                      |  |  |

⇒ Continue to feasibility considerations concerning soft constraints
**Scheme 2: steps for determining feasibility – soft constraints**

<table>
<thead>
<tr>
<th>TYPE OF SOFT CONSTRAINT</th>
<th>TYPE OF SOFT CONSTRAINT</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extent is reaching the preferred outcome free from economic, cultural, institutional and motivational constraints?</td>
<td>The more free, the more feasible</td>
</tr>
<tr>
<td>- At this very moment</td>
<td></td>
</tr>
<tr>
<td>- In the short term</td>
<td></td>
</tr>
<tr>
<td>- In the long term</td>
<td></td>
</tr>
<tr>
<td>In what way is the actor able to reach the preferred outcome through his/her natural human abilities?</td>
<td>Direct diachronic abilities are more feasible than synchronic ones, and synchronic abilities are more feasible than indirect diachronic ones</td>
</tr>
<tr>
<td>- Synchronic abilities</td>
<td></td>
</tr>
<tr>
<td>- Direct diachronic abilities</td>
<td></td>
</tr>
<tr>
<td>- Indirect diachronic abilities</td>
<td></td>
</tr>
<tr>
<td>To what extent is the preferred outcome stable?</td>
<td>The more stable, the more feasible</td>
</tr>
</tbody>
</table>

Before we move on, three important remarks must be made. First, feasibility is not the same as unconditional likelihood (ibid., pp. 817-818). What this implies is that something can be feasible even though it will not happen. If something is feasible, it is not guaranteed that it will take place. Hard constraints can rule out theories by looking at whether the theory is in fact *possible*; soft constraints can be used to support or criticize a theory, possibly in comparison to other theories, by looking at whether it is *probable* that a certain outcome will be obtained. Secondly, *trying* to achieve a certain outcome cannot be deemed feasible or infeasible. People can always try to achieve a certain outcome, whatever the feasibility of that outcome. Thus, the possibility of trying should be seen as independent from the feasibility of an outcome (ibid.). Thirdly, as we have seen, feasibility considerations concerning accessibility can focus on either hard or soft constraints. If there is simply no path towards a preferred outcome, this forms a hard constraint on its feasibility considerations: if there is no road in the first place, we cannot arrive at the town that lies at the other end of the road – the town being the outcome in this analogy. It seems that *stability* constraints can also be both soft and hard. If an outcome “has a value of ‘0’ on stability”, this would mean that an outcome could not even be sustained for one minute after it has been attained. In this case, the stability constraint would be a hard constraint on the feasibility considerations. As soon as we can speak of a certain *level* of stability, the constraints associated with it are soft. To account for this “value of ‘0’ on stability”, I will discuss stability not only as a soft constraint, but also – briefly – as a hard constraint.

---

18 However, an outcome with a value of ‘0’ on stability is probably only a theoretical possibility.
4 THE FEASIBILITY OF HABERMASIAN DISCOURSE ETHICS

In this chapter, I will utilize the account of feasibility developed in the previous chapter to argue that Habermas’s discourse ethics is infeasible. I will look at both Habermas’s theory in general and the discourse rules in his theory separately. It is beyond the scope of this thesis to investigate the feasibility of theories similar to Habermas’s, which makes it difficult to compare the feasibility of Habermasian discourse ethics to other theories. Thus, for this thesis, hard constraints are more important than soft constraints in the sense that the former can show us whether Habermasian discourse ethics can be feasible, whilst the latter show us to what extent it is feasible. First, discourse ethics will be investigated by looking at hard constraints. A discussion of soft constraints will follow. Finally, I will round up the discussion concerning the feasibility of Habermas’s theory and conclude that it is feasible neither regarding hard nor soft constraints.

4.1 Hard constraints concerning discourse ethics
Gilabert and Lawford-Smith (2012, p. 812) assert that we must answer three questions to consider the feasibility of a state of affairs: “Feasible for whom?”, “Feasibility of what?” and “Feasible when and where?” In the present analysis, the answer to the first question will be “(the individuals in) a society that organizes a discourse.” The answer to the second question relates to the preferred outcome of a theory. As can be seen in the short summary above, the preferred outcome of Habermasian discourse ethics is to reach a reasonable agreement. The answer to the third question depends on specific settings and can be situated in almost every temporal and spatial context, so I will not assume anything relating to this question here.

4.1.1 Momentum and inevitability
As can be seen in Scheme 1, the first hard constraint concerns momentum. This constraint is violated when we try to reach a preferred outcome that either will not be or has not been possible, or is inevitable and therefore will definitely be reached. The question that must be asked to determine whether this constraint is present is the following: "Is the preferred outcome either inevitable or impossible in terms of momentum?" Now, can we say that reaching a reasonable agreement is inevitable?

As some of the critics discussed above assert, reaching a reasonable agreement is a very difficult process. As Wellmer (1991) and Rienstra & Hook (2006), among others, have argued, discourse ethics seems to demand very much of agents. Human capacities are not so “divine” that they are suited for reaching a reasonable consensus in the Habermasian sense. Moreover,
the institutions that can be used to reach (or approximate) this ideal set of capacities will contradict discourse ethics. Additionally, from an empirical standpoint, Vivienne Boon (2010) applies discourse ethics to Islamic fundamentalism in order to show that it is not a form of conflict resolution that always works. She argues that discourse ethics only seems to function when participants in a discourse agree to have a deliberative discourse in the first place and share certain basic views on deliberating rationally and on individual autonomy. In short, her conclusion is that if conflicts are to be resolved and there is no willingness to do so in a deliberative manner, discourse ethics does not help us.

The preferred outcome of discourse ethics is clearly not inevitable. But, perhaps more importantly, is it impossible? There are two ways in which we can formulate an answer to this question: practically and conceptually. I will start on the practical level. Aside from the arguments of the above-mentioned authors, Sanders (1997) argues that deliberation in the Habermasian sense leaves power structures and the dominance of certain groups intact. Social hierarchies and oppression on the basis of race, gender or otherwise seem to make participation in a deliberative discourse alienating and subject to the power structures that exist in society. Sanders supports this claim with empirical data on US citizens’ juries. If discourse ethics alienates and leaves oppression and domination intact, a fully reasonable agreement is impossible. Mutz (2008) asserts that deliberative democratic theories as a whole are impossible to measure and exist together, since some of the elements of deliberative theories contradict others in those same theories. Moreover, there is a circularity in the evaluation of the success of deliberation, she argues, because “good deliberation is often defined as deliberation that produces the desired consequences outlined in the theory.” (Mutz, 2008, p. 527) In order to make deliberative theories applicable and measurable, she believes, we should design several “middle-range” theories that contain different elements of a theory, as opposed to trying to evaluate and test it as one, “grand” theory. Still others argue that modern society is too complex for discourse ethics to be possible (Femia, 1996; also see Van Mill, 1996).

Nevertheless, I reckon it is important to note that, as I pointed out earlier, Habermas admits that a fully reasonable consensus is impossible in reality. Moreover, we could utilize multiple series of discourses to create temporary consensi, as Chambers (1993) also suggests. This way, we would be open to changing norms to the values and beliefs of a particular period, but the temporary consensus would also entail an element of continuity. Habermas believes we have to approximate ideal discourse, and with it a fully reasonable consensus, not necessarily achieve it. In addition to this, some studies suggest that Habermasian discourse ethics does steer us towards reasonableness. Barry et al. (2001), for instance, use discourse ethics to identify communicational patterns in doctor-patient interaction. They find that the more communication between doctors and patients concerns the lifeworld of (one of) both, the better and more
humane the treatment of the patients will be. Naturally, this is still far away from reasonable consensus, but it shows that talking about matters that concern agents’ life worlds fosters a more reasonable approach to agents. This shows us the potential benefit of discourse ethics. One of the findings in a case study by Storey and Jacobson (2004, p. 431) is that entertainment-education can stimulate participation and communicative action. Education in the form of entertainment could thus “produce” persons who are (to some extent) suited for participating in a Habermasian discourse. This also emphasizes that the preferred outcome of discourse ethics is at least possible. Therefore, we cannot say that Habermas’s theory is definitely impossible in a practical sense.

Some authors argue that the preferred outcome of Habermasian discourse ethics is impossible to reach in a conceptual sense. Chantal Mouffe (2005, pp. 29-30), for instance, believes that reaching consensus is impossible because conflict is a necessary element of human behaviour. It, and not consensus, characterizes democracy. As she states, “[a] well functioning [sic] democracy calls for a clash of legitimate democratic political positions. [...] If this adversarial configuration is missing, passions cannot be given a democratic outlet and the [conflictual] dynamics of pluralism are hindered.” (Mouffe, 2005, p. 30) Because communication between people is characterized by conflict, reaching a full consensus becomes impossible. Furthermore, as we have seen, trying to reach a consensus seems self-defeating as a process: whenever the consensus is reached, the need for deliberation disappears (Thomassen, 2010, p. 101; Rummens, 2012). However, Habermas admits that a full consensus cannot be reached (Habermas, 1996b, p. 1518; Habermas, 1998a, pp. 365-369). This means that we ultimately strive after achieving an outcome that will never be actually achieved. In this sense, reaching the preferred outcome is impossible, and a hard constraint is violated. If we take reaching a partial consensus as our preferred outcome, the hard constraint might not be violated, but we would also need to state what the preferred level of consensus exactly is. Habermas does not do this, and I believe stating such a thing is very difficult. But how can the (path to the) preferred outcome be feasible if we cannot determine what that outcome is in the first place? Thus, reaching the/a preferred consensus is impossible in both views.

Although discourse ethics is not necessarily impossible on a practical level, we can see that Habermasian discourse ethics has major problems in a conceptual sense. Mainly because of the concerns on the conceptual level, I conclude that the preferred outcome of Habermasian discourse ethics is impossible.

4.1.2 The violation of physical laws (nomological consistency)

The question that must be answered with regard to the second hard constraint I have formulated is: “Does the (path to the) preferred outcome violate physical laws?” I believe that
discourse ethics is not constrained in such a way. One could perhaps assert that, concerning the principle of universalizability (U), it is simply impossible to organize a discourse in which all the people who are affected by the anticipated consequences and side-effects of a moral norm are present. There is no place, at least to my knowledge, that could accommodate all the people affected by the implementation of a norm in a medium-sized nation or even city, for instance. Moreover, it seems too demanding to require all those people to have no other obligations during the period in which the discourse takes place; for what about policemen, doctors, soldiers and other people who are expected to be on call at any time? Should all people just lay down their work and gather, wherever that gathering may take place? To defend Habermas at this point, I must once more appeal to the fact that Habermas wants us to approximate an ideal speech situation and reasonable agreement. I think that, in this case, this implies that not all people who are affected should actually be expected to be present at a discourse. Of course, it would be better if they were, but, as Habermas rightly notes, this is impossible. This way, Habermas provides us with a way in which we can deal with this problem. Furthermore, representative sampling of the citizen population could be used to approximate the deliberative ideal as well as possible (cf. e.g. Smith & Wales, 2000). Also, Weeks (2000) supports Habermas in the sense that he argues relatively large-scale public deliberation is possible, although this is “neither cheap, fast, nor easy.” (Weeks, 2000, p. 371)

4.1.3 A historically fixed starting point

The following question relates to the third hard constraint: “Is the starting point from which the preferred outcome is to be realized situated at a historical place and/or time?” If this question is answered positively, this implies that the point at which the path to the preferred outcome begins (i.e. the “initial situation”) is historical and therefore cannot be influenced and is not relevant. In Habermas’s case, the answer is no. A strength of his critical theory is that it is situated in current society. It is a “part of the object-domain it describes” (Geuss, 1981) and, from this position, provides us with ways in which we can enhance our society, that is, ways in which our self-reflection is induced and we can emancipate and enlighten ourselves. The very hallmark of a critical theory is that it takes as a starting point the current society, often viewed as a society in which individuals live under circumstances of false consciousness. Therefore, discourse ethics is not rendered impossible by this hard constraint.

4.1.4 Logical consistency

The question that explores the presence of the fourth hard constraint is: “Is the path that leads from the current situation to the preferred outcome logically inconsistent?” That is, do the concepts and expectations that are part of the expected path to the preferred outcome (not)
share a logical framework? One (indirect) response to this question is given by Cristina Lafont (2006). She defends deliberative democratic theory against the claim that it is logically inconsistent because the notions of democracy and deliberation are incompatible. She asserts we cannot defend deliberative democracy if deliberation is (partially) reduced to democracy, or vice versa (as is often done, she argues), because both have a central place in the theory. Lafont (2006, pp. 7-8) believes that we can combine the two notions consistently. She thinks that the deliberative democratic ideal does not require citizens to give up the deliberative goal for the other, or vice versa, and that there is no logical inconsistency between the two notions (ibid., pp. 20-23).

Jane Mansbridge (2007) asserts that some assumptions Habermasian discourse ethics makes are disputable. For example, she writes that Habermas finds that “only reasoning on the common good can generate legitimate law.” (Mansbridge, 2007, p. 261) Mansbridge points out that we do not use “reason” in discourse, but “reason-giving”, the latter of which implies multiple kinds of reasons. She believes, based on works by Marcus (2002), Nussbaum (1995) and Rorty (1985), that Habermas wrongfully neglects the important roles that emotions, bargaining and negotiation play in reason-giving besides reason as such. Mansbridge asserts that “emotional discourse intertwines inextricably with rational discourse in deliberation.” (Mansbridge, 2007, p. 262; also see Warren, 1993) An author who agrees with Mansbridge is Iris Marion Young. Young argues that the model of deliberative democracy tends to have an exclusionary bias: it “restrict[s] democratic discussion to argument” (Young, 1996, p. 122; also see Young, 2000). This way, individuals who like to argue with others or who are good at arguing are privileged and individuals who do/are not receive a lower level of privilege in the discourse. There are also other ways in which people can communicate with the aim to reach understanding (greeting, rhetoric and storytelling are aspects of communication Young discusses). Speech does not only imply rationality or argumentation, but also emotions and rhetoric. Lynn Sanders (1997) also voices this point of critique and believes we should not only value argument in deliberation, but also testimony. Young also asserts that the assumption of deliberative democratic theories that “unity is either a starting point or goal of democratic discussion” is potentially exclusionary as well (Young, 1996, p. 122). This is because striving after a “common good” or unity may privilege the classes or groups in society that are already dominant (ibid., pp. 125-126). This is what Sanders (1997) also contends.

Although I must say that Habermas does not explicitly provide a defence against the criticism on the force of argument as the sole standard in discourse, Wright (2004) and Keller (2008) do so, as I mentioned earlier. Recall that Wright argues that Habermas allows for feelings within and without discourses, because the boundary between ethical and normative expectations is not fixed in discourse ethics, and people themselves can therefore determine
where to draw the boundary (Wright, 2004, pp. 58-63). Keller (2008, pp. 177-178) argues that discourse ethics presupposes forms of receptivity, assertiveness and reciprocity, and that social and emotional capacities are crucial for developing these personal qualities. However, it can still be said that these capacities only count for the process of deliberation in discourse ethics. That is, they help us to understand each other better and enhance our social abilities, but they are not seen as relevant in the actual discussion of the validity of moral norms. Norms are only seen as valid if the arguments that are given are acceptable. So discourse perhaps requires some emotional capacities, but does not allow for these capacities to influence the actual “decision-making process” (i.e. the determination of the validity of moral norms). Emotions and empathy can certainly make it easier to understand other participants, but are not part of the standard that is used to assess the validity of a moral norm. Therefore, argumentative power is still the sole standard for the validity of norms, which privileges individuals who like to argue or are good at arguing over others.

A third logical inconsistency in discourse ethics is presented by Chambers (2004). She believes that some authors are right in pointing out that “the “glare” of publicity makes it difficult to argue publicly” (Chambers, 2004, p. 389). In other words, individuals find it harder to deliberate in the presence of others than in an isolated setting. This would mean that deliberation in secret would result in “more publicizable outcomes”, and that the outcome of deliberation will not be as positive as deliberative democratic theorists think. Nevertheless, Chambers argues that we can overcome this problem by distinguishing three forms of reason: private reason, which is not general enough to appeal to all affected; plebiscitory [sic] reason, which is too poorly argued, manipulative or shallow to serve as justification of public policy; and public reason, which is suited for the justification of public policy. She believes that by containing plebiscitary reason and promoting public reason, the deliberative democratic ideal becomes more attainable. Another solution to this inconsistency is offered by Sanders (1997). She argues that we should supplement the public discourse with individual testimony, so that people are not constrained by the “publicness” of the deliberative setting.

I have only discussed three potential incoherencies in Habermas’s theory, but as we have seen, some arguments against Habermas are met by several defences of Habermas’s theory. In my opinion, it is right to conclude that in spite of these responses, there are many consistency problems with discourse ethics. This hard constraint is therefore violated.

4.1.5 Biological/psychological and natural human ability constraints

The questions relating to biological and natural human ability constraints are the following: “Is there a biological or psychological constraint on the (path to the) preferred outcome?” and “Is/will the actor that is to bring the preferred outcome about (be) unable – in both a direct and indirect
sense, in both the present and the future – to bring the outcome about?" I discuss these questions together in this section, because they strongly relate to each other as hard constraints.

Some authors have argued that human beings are incapable of possessing or using such rational faculties or intellect that Habermas assumes them to have\(^\text{19}\) (cf. Wellmer, 1991; Rienstra & Hook, 2006). Furthermore, Shawn Rosenberg (2007) utilizes empirical data to investigate the behaviour of individuals in several discourses. He concludes that people do think beyond their narrow self-interests, but do not act in the reasonable, rational or critically reflective way they are assumed to do (Rosenberg, 2007, p. 157). This, he contends, is problematic for discourse, because people are only limitedly self-reflective or socially critical; because this maintains hierarchical and social divisions that already exist; and because people are unable to understand and address differences in social norms in the way a multicultural society expects them to be (ibid.). Rosenberg believes that the lack of reasonable, rational or critically reflective behaviour is mainly caused by humans' natural inability to act in the way discourse ethics asks of them. This forms a hard constraint on discourse ethics, because the (original) preferred outcome of reasonable consensus cannot be reached.

Warren (1993), as well as Young (1996), Sanders (1997) and Mansbridge (2007), point to another problem: Habermas only pays attention to language as the mode of communication in a discourse, and therefore neglects non-lingual or subconscious parts of communication. Although language is a very important part of the communication between people, human beings simply do not merely communicate through language, but also through body language, empathy etc. The result is that Habermas "sees non-lingual communication as illegitimate", as not being a part of people's needs and interests for the self. If this is so, people will never become, as selves, the reasonable and critically reflective persons they want to be. Moreover, there will be a "layer" of argument that is not accounted for: agreement or conflict can result from body language as quickly as from spoken or written language. Therefore, Habermas's theory has a limited scope of communication. It sees human communication as simpler than it actually is. Furthermore, several authors would say that power structures cannot be eliminated from deliberation, and thus that coming to a (full) reasonable agreement is impossible (Young, 1996; Sanders, 1997; also see Morriss, 2002; Lukes, 2005). Moreover, it is part of human nature to be influenced by group structures and others' behaviour (Janis, 1982; Young, 1996), which makes it even more complicated to "forget" power structures or compensate for others' superior eloquence, persuasive power etc. and can result into so-called cognitive dissonance reduction\(^\text{20}\).

\(^{19}\)This relates to human beings' "bounded rationality". Also see Fearon, 1998, p. 49; Smith & Wales, 2000, pp. 53-54.

\(^{20}\)This is, shortly put, emphasizing the benefits of a certain option, policy, value etc. and thereby reducing the tension between the proposed option, policy or value and a person's personal beliefs and values.
Nevertheless, Ratner (2008) finds that people sometimes actually approximate the ideal speech situation we should strive after. Moreover, he finds that if we look at the future, the image becomes even more positive, because people actually seem to steadily learn and develop the capacities that are needed in a reasonable discourse (Ratner, 2008, pp. 160-161). This potential is supported by Walsh (2007, p. 46), who argues that “including dialogue in deliberative systems has the potential to improve intercultural connections by enabling participants to construct a balance between unity and diversity.” In a similar vein, Jason Barabas (2004) argues that deliberation potentially has positive effects on participants’ reasonableness. He finds that participants learn readily, encounter and learn to understand different perspectives and soften views they hold strongly. Deliberation alters and soften people’s opinions and increases their knowledge. However, this does depend upon the diversity and the quality of the messages people use, as well as upon participants’ willingness to be open to others’ views and opinions.

We can infer from these studies that the findings are predominantly negative. People cannot be fully reasonable, rational and self-reflective, which makes reaching a reasonable consensus impossible. Still, what some of the above-mentioned studies suggest is that people can learn to increasingly behave like “reasonable” people – at least some people. This means that the constraint mainly holds for now, but can partially be overcome later on, and this only counts for some people. In spite of this potential to enhance individuals’ “deliberative skills”, individuals will never be able to reach (or approximate) a state of full reasonableness. Moreover, only a restricted number of people have the chance to develop their skills, which hardly seems inclusive and equitable vis-à-vis the people who do not have this possibility. Thus, individuals’ lack of psychological and biological capacities forms a hard constraint on the preferred outcome of reasonable consensus: it renders it impossible.

4.1.6 Stability

When looking at stability, we must ask: “Does the preferred outcome completely lack stability?” What is important is that this question is about the complete lack of feasibility. I believe there is no problem for Habermas here, because I think we can assume that if people take the trouble to participate in a discourse and eventually reach an agreement that is at least to a great extent reasonable, they will not change their opinions immediately after the discourse. And if they do, they have probably been insincere, which would violate Habermasian discourse rules. Naturally, their values and opinions could change over time, perhaps within a few weeks; but this is no problem for Habermas, because he argues that in that case, we need to organize a new discourse to reach a new agreement that is acceptable in the light of the changed values and opinions of certain people. Furthermore, discourse participants will know that, as soon as a moral norm is
justified (at least to a great extent), it will probably be implemented through certain policies or laws. Thus, first, I believe they will think seriously about what they are deliberating about – this even concerns one of the discourse rules of Habermas’s theory – and secondly, the very fact that the valid norms will probably be implemented forms a form of stability for those norms as such. Therefore, I believe that discourse ethics is not completely infeasible with regard to stability constraints.

4.2 Soft constraints concerning discourse ethics

I will now look at the extent to which Habermasian discourse ethics is feasible – soft constraints. Here, it is assumed that the actors involved and the preferred outcome are the same as in the discussion of hard constraints above. I will look at each soft constraint in turn.

4.2.1 Economic, cultural, institutional and motivational constraints

The question that must be asked here is: “To what extent is reaching the preferred outcome free from economic, cultural, institutional and motivational constraints?” This general question is subdivided into three more specific questions (see Scheme 2 above), which consider these constraints at this very moment, in the short term and in the long term. I will discuss these three questions here.

In my opinion, one of the most prominent constraints for discourse ethics is the (partial) colonization of the lifeworld by the system in many (western) societies. In an economic sense, the lifeworld lacks prominence and importance in current societies, because the system, driven by instrumental rationality, dominates the lifeworld. This does not only mean that the presence or importance of the economy is too large in current societies; it also implies that communication in the lifeworld is, to a large extent, based on conceptions of instrumental rather than communicative rationality. How people see each other, what values people in a society share, what “happiness” is; all these matters are strongly influenced by a capitalist, instrumental Weltanschauung. For instance, TV and Internet commercials stimulate us to buy things we do not actually need, or people often define who they are by talking about what they have or what their parents do. In this sense, “successful” persons are often those who have the newest gadgets, expensive clothes and a monthly salary that is much higher than what they actually need to lead a “decent” human life. Moreover, the institutions in these societies are driven by these instrumental goals and policies as well, because they have become a “way of life” in society as a whole. For instance, art galleries, cabaret shows and traditional literature are often denied subsidies or are often seen as “unnecessary” or “unimportant” matters because they do not provide us with money or property. They predominantly provide us with intangible forms of
happiness: knowledge, laughter, emotional satisfaction etc. Long-lasting institutions must change with society, because their chances of survival are not high if they hold views that are radically different from the main views in society. Furthermore, new institutions will reflect the views held in society, which are, of course, already heavily influenced by instrumentally rational aims. Now, although Habermas does not believe that instrumental rationality is wrong per se, he contends that it is too dominant in present societies. It should be limited to the system only, and the lifeworld must be driven by communicative rationality. The system will be influenced by the communicatively rational discourses organized by people in their lifeworlds, but is not expected to abandon the instrumentally rational goals and procedures it has.

What this boils down to is that present societies are in many ways radically different from the “ideal society” that discourse ethics aims to attain. Habermas argues that we can overcome this situation by participating in a discourse aimed at reaching a reasonable consensus, but I believe it is very difficult to start a reasonable discourse in a society mainly driven by instrumentally rational goals and procedures in the first place. In other words, I believe that the self-reflective capacity he wants people to obtain through deliberation is unachievable – at least not in the way Habermas argues that it is. As we have seen above, some studies support my argument here (Wellmer, 1991; Rienstra & Hook, 2007; Rosenberg, 2007) and some studies find that we can enhance people’s ability to participate in a discourse, albeit only up to a certain level (Barabas, 2004; Ratner, 2008). Nevertheless, even if people are able to enhance their abilities in such a way, the path towards a high “level” of reasonableness will be long and difficult, I believe. In my view, the radical differences between the current state of affairs and the preferred outcome pose a serious soft constraint on deliberation: it renders attaining the preferred outcome less probable.

There is another problem concerning these two forms of rationality, described by Kenneth Avio (1996). Avio has written several articles about the practicability of discourse ethics in the first sense. In a summarizing article (1996), he discusses his work on comparing Buchanan’s to Habermas’s theory and applying both theories to the Aboriginal case. He looks at what a discourse-ethical perspective would entail for the Aboriginals. He concludes that Aboriginal sovereignty could be restored if an ongoing discourse about implementation of the norm of cultural equality breaks down: if the discourse breaks down, the parties in that discourse lose their sense of cultural identity, Avio argues, and thus, they will be back at the point where they started, providing the Aboriginals with an opportunity to claim their sovereignty. Avio uses discourse ethics to show that Habermas forgets the fact that a switch

---

21 Also see Habermas’s explication of his “two-track model of democracy”: Habermas, 1998b, pp. 239-264.
22 For a practical, but limited discussion of this problem, see Grundy, 1994. Shabani (2010) also recognizes this problem in his case study, but asserts that it can be overcome by promoting tolerance and non-violence in that specific case. However, I wonder whether Shabani’s case is generalizable to other cases.
from reasonable expectations to rational expectations must take place in order to make norms work for the parties in a discourse, which always have a set of opportunity costs and are in circumstances of scarcity. This means that on the one hand, participants in a discourse must think beyond their narrow self-interests and act/think in a reasonable way; but on the other hand, the consequences of validated moral norms must be interpreted in an instrumentally rational way as well. For an individual must know what the validation of a moral norms implies for his personal opportunity costs and circumstances of scarcity. In order to know this, the instrumental implications of a norm’s validation must be laid out as well, because opportunity costs and scarcity are instrumental constraints. Avio thinks that if a discourse of implementation is added to discourse ethics to create a compromise between parties, parties will really be able to “start over again” – which would be fortunate for the Aboriginals. The main point of this argument is that many people cannot only relate to reasonableness and will need to know what they can expect of a policy or norms in an instrumentally rational sense. This effectively alters the aim of the discourse, because people will try to attain a compromise instead of a reasonable consensus. This, again, shows how hard it will be for people to abandon their instrumentally rationally influenced views in a Habermasian discourse.

In another article (1997), Avio indirectly points towards a second problem. He establishes that “[f]or Habermas, the task [of the appropriate design of social institutions] is to design institutions that foster undistorted communication” (Avio, 1997, p. 551). This kind of communication must be in accordance with the “universal and unavoidable presuppositions of action oriented to understanding” (Habermas, 1992, p. 191). His main point is that institutions must shape an environment that is an approximation of the ideal speech situation and minimizes the influence of people’s narrow self-interests. However, I reckon it is very improbable that institutions can provide such an environment if the institutions themselves, heavily influenced by instrumental rationality, do not have an incentive to do so. The point of Habermas’s theory here is that communicative rationality must become dominant in the lifeworld and subsequently influence policies and laws in the system. Therefore, we must somehow start with the people themselves, who can change the institutions and values that are present in the current lifeworld. And precisely this seems to be a long and difficult path.

Nevertheless, there are authors that see comparative benefits for discourse ethics. Simone Chambers (1998), for instance, applies Lockean contractarian thinking and Habermasian discourse ethics to constitution-making in Canada. The aim of her article is to investigate which theory is the most realistic in terms of constitution making. She argues in favour of discourse ethics and rejects Lockean contractarianism because “contracts cannot accommodate deep diversity” in society (Chambers, 1998, p. 144) and discourse ethics can, because agreement is/can be maintained over time: discourses can be initiated if norms are not justifiable anymore,
whilst a traditional political contract does not provide suchlike opportunities. A contract is made at a particular point in time and cannot be revised. Moreover, people will expect that the contract contains the unitary vision of the people, and that this unitary vision is also their own. In societies characterized by deep diversity, there is no unitary vision, because there is no homogeneous “people”. Still, there are two problems with this account: first, exactly because discourse ethics can accommodate diversity and multiculturalism better than Lockean contractarianism, it will become even harder to come to a reasonable agreement. The more people with different views have to participate in a discourse, the harder it becomes to reach an agreement, I believe. Secondly, Chambers has compared discourse ethics and Lockean contractarianism in only one case. It could be that her findings are not generalizable to other nations or sub-, inter- or supranational bodies or groups (for example because of the fact that Canada is a multination state as well as a polyethnic state; cf. Kymlicka, 1996, pp. 153-159).

In their review of the literature on the empirical investigation of deliberative democracy, Delli Carpini et al. (2004) state that the problems mentioned above are legitimate problems, but they sketch a more nuanced image. In their opinion, a lot of empirical research on political deliberation “provides a good deal of indirect support for the democratic potential of deliberation but also suggests that this potential is highly context dependent [sic] and rife with opportunities for going awry.” (Delli Carpini et al., 2004, p. 328) The impact of deliberation “is complex and context dependent [sic] and does not assure either citizen satisfaction or government responsiveness.” (ibid., p. 332) Experimental research on political deliberation “paints an equally if not more complex picture” (ibid.), as does case-study research and deliberative-polling research. Although the evidence used in these studies is often “partial and inconsistent”, the authors conclude that “the impact of deliberation and other forms of discursive politics is highly context dependent [sic].” (ibid., p. 336)

There is a constraint I have not yet touched upon here: motivation. As I stated in the previous chapter, a preferred outcome can become more feasible if people become more motivated to reach the outcome or be involved in reaching the outcome. Now, will people be motivated to participate in a discourse? I reckon that, precisely because people live in a society in which the system colonizes the lifeworld, and because people have got used to this colonization and its consequences, they will, at least in the first instance, not be motivated to change this23. First, we have seen that human beings’ capacity to be, or become, reasonable and self-reflective is limited. This impedes the process in which they come to realize that they live, and have lived, under conditions of “self-imposed coercion” and “self-delusion” (Geuss, 1981, p. 58). If this process of enlightenment and emancipation cannot be completed, the preferred

---

23 This assertion is supported by very low turnout rates for citizens’ juries or other forms of deliberation. See, for instance, Berry et al., 1993; Eliasoph, 1998; Hibbing & Theiss-Morse, 2002; Delli Carpini et al., 2004.
agreement will not be reached and people will not be motivated to try and reach it. Essentially, this means that exactly because of the colonization of the lifeworld, reaching a reasonable agreement becomes highly improbable. In this sense, Habermas’s social theory works against his moral theory of discourse ethics.

Secondly, as I also showed earlier, power structures and hierarchical relations cannot be completely kept out of a discourse\(^{24}\). Discourse participants will be influenced, and probably motivated, by factors such as other participants’ superiority (in any sense); loyalty to a certain group among the participants; and “belonging to the normal people” or, as Foucault would say, they are subject to a “normalizing gaze” (Foucault, 1977)\(^{25}\). Thus, people would be motivated by many other goals than merely the communicatively rational ones Habermas expects them to have. Perhaps these other goals would motivate them to participate in a discourse as such, but not always to strive after the ideal of a reasonable consensus as the goal of the discourse.

Moreover, if only very few norms will actually be seen as valid (cf. Finlayson, 2005, pp. 80-81, 87-88), people might lose their motivation to carry on deliberating. For what is the use of putting time, energy and money into a discourse that provides us with only a very small number of acceptable norms? Of course, time, energy and money problems would not be the most important concerns in terms of communicative rationality, but I cannot be persuaded to think that people will abandon these problems when they participate in a discourse. I reckon these matters will be important to people to such an extent that only accepting a few moral norms seems to be a lousy result. Related to this is that, to become motivated to participate in a discourse in the first place, people must be convinced by good arguments – at least that is what Habermas would say, I think. But how can arguments be “full” and convincing enough (in the communicatively rational sense) that people will be motivated to participate if those arguments are created by deliberation? In other words, there can be no communicatively rational arguments for participation if people are not motivated, but without these arguments, there is a higher chance that people are not motivated to participate. A painful circularity becomes visible here. The arguments in favour of deliberation could be developed through education, but this

\(^{24}\) Although Mendelberg and Karpowitz (2007) argue that some designs of deliberation amongst citizens are more susceptible to power structures and group patterns than others, I am still convinced that these factors cannot be phased out in any design whatsoever.

\(^{25}\) This is also important when looking at culture or socio-economic class. If people want to favour or belong to a certain cultural or economic group in society, this will impede the process of reaching a reasonable agreement.
would also be problematic, because education is always based on a certain (cultural) conception of the world; a bias.

To nuance the view, I turn to Delli Carpini et al. (2004) once again. They argue that, concerning motivational constraints, if people are motivated to “think in depth about the essential merits of a message” (Mendelberg, 2002, p. 166), it is more likely that they will participate in deliberative discussions and to create arguments that are valid (Delli Carpini et al., 2004, p. 326; also see Cacioppo & Petty, 1982; Cacioppo et al., 1996; Shestowsky et al., 1998). However, and importantly, these people are also “more resistant to the views and arguments of others” (Delli Carpini et al., 2004, p. 326; also see Petty et al., 1995).

Motivational factors constrain the feasibility of discourse ethics in such a way that both the goal of a discourse and the procedure of the discourse itself become increasingly difficult to realize/strive after. The situation current (western) societies are in, viz. a situation in which the lifeworld is colonized by the system, poses very important feasibility problems for Habermasian discourse ethics. Although it may be argued that these problems can be partially solved in the long term, I believe they will always be present, and will always impede the preferred outcome of discourse ethics. Therefore, I conclude this section by saying that realizing the (path to the) preferred outcome in discourse ethics is fairly improbable.

4.2.2 Natural human abilities

I have already pointed out that, at this very moment as well as in the short term, the approximation of a reasonable consensus through democratic discourse is difficult and context-dependent. It seems that people have no synchronic ability to reach the preferred outcome; that is, they are not able to reach the preferred outcome now. Neither are they able to reach the preferred outcome later without the requirement of other actions. That is, they do not have a direct diachronic ability to do so, because some actions must be performed for people to become able to participate self-reflectively and reasonably in a discourse, and approximating a reasonable consensus. For instance, people must learn to acquire a better understanding of others’ views and soften their own views (Barabas, 2004); people must have the practical possibility to deliberative in a discourse-theoretical manner in some way; people must be able to resist, at least to a large extent, the pressures of the social, economic and cultural groups they are part of in society outside of the discourse; and people must be motivated – by themselves or others – to take part in a discourse in the first place. Therefore, (the path to) reaching the preferred outcome can be best seen as an indirect diachronic ability; that is, the outcome can only be reached on a later point in time, and reaching the outcome presupposes other actions.
4.2.3 Stability

The question that relates to the last soft constraint is: “To what extent is the preferred outcome stable?” As I have argued above, I think if people commit themselves to participating in a discourse and eventually reaching an agreement that is, at least to a great extent, reasonable, they will not change their beliefs and values directly after the discourse. The agreement that is reached in a discourse is there to validate norms and, indirectly, to implement policies and design laws. Policies and laws are mainly focussed on a period that spans several years, so the norms that they are based on must be, too. Moreover, the time, energy and money that is needed for the organization of a discourse keeps people from constantly changing their opinions and “forces” people to think carefully about what they actually want. On the basis of these considerations, I think that whenever a reasonable consensus is approximated, that outcome will be stable enough. Of course, politics is always concerned with designing new policies, but the norms that are the basis for the framework in which these policies are made do not have to change continuously.

4.3 Constraints for discourse rules specifically

I will now discuss the kinds of constraints that concern Habermas’s discourse rules specifically. The discourse rules form an important part of Habermas’s theory. In fact, they prescribe how people should act in a discourse, and what constitutes reasonableness. I assert that changing the structure of the discourse rules, or some rules specifically, is one of the most promising ways of increasing the feasibility of Habermasian discourse ethics, because the set of rules forms one of the most important and clearest elements of Habermas’s theory. I will now briefly touch upon each level of discourse rules in turn.

4.3.1 The logical level

The first rule on the logical level is that nobody is allowed to contradict himself. I believe that this rule is worthy of striving after, but not realistic. People often do not actually know that they contradict themselves. This is not necessarily a problem, but it becomes problematic if discourse participants are unable to discover and/or voice contradictions in other participants’ assertions: should we (wrongly) expect everybody to have such magnificent cognitive and argumentative capacities that always point our self-contradictions out to us? Moreover, people sometimes do not want to acknowledge that they are mistaken26. Especially because people need time to acquire the abilities suitable for discourse (cf. e.g. Barabas, 2004) and because people will not

---

26 This argument relates to the sincerity of discourse participants (rule 2.1), but because the rules on the logical level are discussed first, the argument is discussed here.
start the discourse in a fully self-reflective and reasonable way, I believe they will regularly be unable to point out their own mistakes, or even admit them when they know they are mistaken.

The second rule states that “every speaker who applies predicate $F$ to object $A$ must be prepared to apply $F$ to all other objects resembling $A$ in all relevant aspects” (Habermas, 1990, p. 87). I believe this rule can work out relatively well, because this is something participants can agree on and write down. That way, if somebody uses a different predicate for the same object, others can remind him/her of the fact that he is using the wrong predicate. A discussion could follow about whether the predicate must be changed or not, and eventually, a certain predicate for the object in question will be determined. This gives people stability and, in spite of the amount of time it will cost to write down all (relevant) predicates, I believe it does not go beyond communicational or psychological human abilities. Neither does it assume things that are simply impossible (e.g. logical or nomological inconsistency) nor is it heavily constrained by economic, cultural or institutional factors. It is something every participant can do, or at least be expected to be able to do.

Thirdly, different speakers are not allowed to use the same expression with a different meaning. Although I believe there are no hard constraints that this rule poses – it is not impossible – the probability that all participants stick to this rule seems rather low. For the meaning of words and expressions is heavily influenced by culture (cf. Hanson, 1975; Tennekes, 1992). Perhaps this problem will not arise in a discourse in a small community, but if we organize a national or provincial discourse, there will be cultural sub-groups, immigrant groups, people from different socio-economic classes etc. Assuming that everybody speaks only one language in the discourse, the interpretation of words (and body language) is so diverse among people that full agreement on the meaning of words will, in my opinion, not even be approximated.

4.3.2 The procedural level
The first rule on this level is that “every speaker may assert only what he really believes” (Habermas, 1990, p. 88). Just like my conviction that people often do not know that they contradict themselves, I believe that people often do not know what they really believe. This, too, relates to the social environment one is brought up in, because all norms and values a person has are shaped, or at least influenced, by her social surroundings and culture. Moreover, groupthink issues and power structures (cf. Janis, 1982; Morriss, 2002; Lukes, 2005), which, as we have seen, partially persist in the discourse, can shape beliefs and opinions in such a way that people assert certain convictions in order to favour a group or a dominant force in the discourse – sometimes even without consciously knowing. Arguably, if people do not know what they believe, they will develop their beliefs through the deliberative procedure itself, by listening and
talking to other participants. However, problems concerning “groupthink” patterns and power structures will persist and will perhaps become graver as well: if a participant’s beliefs are formed through deliberating with others, this means that others’ convictions largely influence his/her beliefs, which can be seen as a form of power difference.

As a second rule, it is stated that the disputation of a norm or proposition must be accompanied by a reason for wanting to do so. I do not think that this is impossible, nor highly improbable. We could simply agree that a reason has to be given for every refutation of a norm, and I am confident that most people would agree to this. For in my opinion, people will soon understand that if they themselves have to give a reason in such a case – which they can find frustrating – other people must also do so. I think this “reciprocity” will instil in them a willingness to give reasons for their refutations themselves, even if they actually do not want to.

The bigger problem concerns the question what kind of reason is to be given. People give all sorts of reasons all the time, but many of these are neither compelling nor true, reasonable or rational. If providing any reason is enough, reaching the preferred outcome will probably be impeded, because we will have to determine what kind of reasons are acceptable, which takes time and costs energy and money. Although Habermas believes this is necessary, this poses a soft constraint for the discourse rule, because the time, money and energy spent on the process make the deliberative enterprise longer and more troublesome than would be the case if this discourse rule were not a part of Habermas’s theory. However, if we do not judge these reasons on their reasonableness, truth etc., we will not be able to discern wrong or bad reasons from good or better ones. And precisely this discernment must be present to let the “better argument” prevail in deliberation.

4.3.3 The process level
“Every subject with the competence to speak and act is allowed to take part in a discourse” (Habermas, 1990, p. 89) is the first discourse rule Habermas lists on this level. I reckon that this rule is too vague to be feasible, and also discriminating. What is the “competence to speak and act” precisely? Is speaking merely the ability to produce sounds? In that case, young children, (mildly) mentally and physically handicapped people and perhaps even animals should be included in the discourse. And is acting simply the ability to perform an action? I do not believe that Habermas would favour such an interpretation. Nevertheless, he does not provide us with a convincing answer. Abiding to this discourse rule means drawing a line between people “fit” for discourse and those who are not, or between people that “matter” in a discourse and people who do not. It could be argued that people who are not “fit” for deliberation can be represented by people who are. Although I assert that this is probably the best way in which this problem can be coped with, it still gives the people excluded from the discourse a lower moral standing than the
other participants, exactly because the “unfit” people are represented and normal, “fit” people are not. If representation is to be carried out in an equitable way, more people than just the “unfit” ones should be represented. Aside from this, speaking and acting is the criterion, but the ability to feel, for instance, is left out (also see the earlier comments based on Warren, 1993; Young, 1996; Sanders, 1997; and Mansbridge, 2007). Habermas seems to want to include participants who possess predominantly masculine abilities. Inclusion is very important for Habermas, because, in his view, we should dialogically come to validate (or refute) moral norms. This means that we must include all sorts of people into our conversation to achieve a substantial level of universality. Full inclusion cannot be achieved, but representation of people “unfit” for deliberation can be regarded as a form of a solution. However, as was mentioned before, representation should best be done across the whole population. Moreover, the discourse rule is too vague to draw a (justifiable) line between participants and non-participants, which is a problem in itself.

Secondly, there are three rules under one point, stating that everybody may question any assertion whatsoever; everybody may introduce any assertion whatsoever; and everybody may express his attitudes, desires and needs. I think these rules are very valuable for creating a (relatively) free and high-quality discourse. As Habermas argues, discourse participants will decide which assertions are relevant and which are not through the deliberative procedure itself. This safeguards against participants making claims that are “irrelevant” to the discourse. Therefore, expressions that are not deemed relevant to the discourse will soon disappear from the discourse. I believe the basic thought of these three sub-rules is right. Moreover, I believe that determining the relevance of expressions can be established through deliberation. These rules have a high level of feasibility.

Rule number three says that nobody may be prevented, by internal or external coercion, from exercising his rights as laid down. In my opinion, it is right to strive after this goal, but it will never be accomplished or approximated to a great extent. As I have already mentioned multiple times, power structures, group loyalty, cognitive dissonance reduction and hierarchy persist in a Habermasian discourse. Moreover, some persons will have a greater talent for public speaking or convincing others than their fellow participants. I believe such talents cannot be fully compensated, which necessarily creates inequalities amongst participants (cf. Young, 1996). Moreover, as I have showed above, it is highly improbable that all participants in a discourse attain a high level of reasonableness. I think this rule is highly infeasible because Habermas does not distinguish multiple kinds of coercion or inequality. Being forced to cease talking is something else than becoming silent because you believe that is better for the group you think you belong to. People live under conditions of false consciousness and this will continue to pose difficulties for reaching the preferred outcome. However, as Geuss (1981, p. 58)
states, the assumption that Habermas makes is that “[t]he ‘unfree existence’ from which the agents [...] suffer is a form of self-imposed coercion; their false consciousness is a form of self-delusion” and “the coercion [...] is one whose ‘power’ or ‘objectivity’ derives only from the fact that the agents do not realize that it is self-imposed”. Or, as I see it, there is inequality and people suffer, but they do not know why, or they think it is justified that this is so. The suffering people believe that this is the way it is for them. They know nothing else. Therefore, they will act on the basis of what they know – as every human being does. They do not see their behaviour as internal coercion. The dominant group in society, however, will try to maintain this order through external coercion. They consciously coerce others. In my opinion, this element of consciousness shows that external coercion is often less excusable than internal coercion. Therefore, I think we must generally regard external coercion in society, but also in discourse, as more damaging than internal coercion. Moreover, some forms of coercion are graver than others in general, I contend. Rule (3.3) does not make a distinction between levels of graveness of coercion. I believe, for instance, that physical violence and grave mental and psychological abuse are more severe forms of harm than mild mental abuse, offensive behaviour and disgust towards primary objects (cf. Nussbaum, 2010). Habermas does not account for this difference in rule (3.3). Nevertheless, I reckon we should prioritize graver harms over lesser harms, precisely because the former are less excusable than the latter. I will briefly return to this point at the end of the next chapter.

4.4 Discourse ethics and feasibility

In this chapter, I argued that Habermasian discourse ethics is not fully constrained by all of the hard constraints I have formulated. However, when we look at degrees of feasibility, or probability, Habermas’s theory does not seem very feasible. It seems highly improbable that people are capable of participating in a discourse in the way that Habermas expects them to. Furthermore, reaching, or even approximating, a reasonable consensus is very intensive and will take a long time. In spite of this, the preferred outcome of discourse ethics seems to be rather stable. Concerning the discourse rules, the general conclusion is that most of them carry with them valuable and right ideals, but that these ideals will never be reached or approximated, because the rules are too vaguely formulated or too demanding for human beings. It would be interesting to know how we can alter the discourse rules to increase the feasibility of discourse ethics. I will touch upon this in the next chapter.

---

27 There are exceptions to this, but I will not go into them in this thesis.
5 A NEW FEASIBILITY FRAMEWORK

In this chapter, a new framework for feasibility will be presented. The main concern here is to show that the conception of feasibility I formulated earlier can be best seen as a part of a broader feasibility framework. The conception of feasibility that was used in the previous chapter is incomplete, because it does not account for all important facets of feasibility. Moreover, it is shown how this framework can be applied to Habermasian discourse ethics by using it to rank the discourse rules Habermas has formulated. This way of applying the new framework, I argue, can be used for other parts of Habermas’s theory, or for other theories in general. First, a conceptualization of the new feasibility framework will be presented. Subsequently, this framework will be applied to the discourse rules. The chapter ends with several alterations of the discourse rules that I believe would make Habermas’s enterprise more worthwhile.

5.1 A broader notion of feasibility

Many of the feasibility problems with Habermas’s theory are connected to the discourse rules he formulates. For instance, “Every speaker may assert only what he really believes” (2.1) (Habermas, 1990, p. 87) relates back to the problem of false consciousness Habermas himself describes: some people will not be able to express their “real” needs – insofar as these actually exist – and will therefore only express false or corrupt ones. A second example concerns rule (3.1): “Every subject with the competence to speak and act is allowed to take part in a discourse” (ibid., p. 88). A problem connected to this rule is that not only the argument should count in a discourse, but also rhetoric, emotion, testimony and so forth. Moreover, including every individual with the competence to speak and act amounts to the exclusion of every individual who does not have the (complete) competence to do so.

Because the rules are of central importance to discourse ethics, feasibility problems concerning them pose serious constraints on Habermas’s theory. I do not deny that these problems cannot be fully solved. Nor do I deny that Habermas himself has admitted that the preferred outcome of his theory cannot be fully achieved. Furthermore, I contend that every discourse rule Habermas presents in his work is important to the enterprise of deliberation. Importantly, Habermas asserts that accepting these rules is necessary for participating in a discourse and attaining a reasonable agreement. If an individual denies the importance of any of these rules, this results in a performative contradiction: the denial of one or more “unavoidable” presuppositions contradicts the practice of the discourse itself (ibid., pp. 82-86).
Nevertheless, as was argued earlier, the preferred outcome cannot be fully reached, nor, perhaps, approximated to a large extent. Now, what does this imply for the importance of the discourse rules? I do not believe they remain to be necessary presuppositions for the discursive enterprise, because people cannot be in full obedience to the discourse rules. Therefore, both the preferred outcome and the discourse rules necessarily become a matter of degree: the discourse rules must be accepted (or obeyed) to a certain extent in order to reach agreement to a certain extent. Thus, I contend we ought not to regard the discourse rules as dichotomous variables, but as ordinal ones: the boundaries between levels of acceptance of (or obedience to) the discourse rules cannot be drawn clearly, but we can distinguish several levels of acceptance themselves. On the basis of these considerations, we could say that the larger the extent to which a rule is accepted or obeyed is, the larger the extent to which the preferred outcome is approximated becomes. It should be noted that some discourse rules can possibly be of a greater importance than others. I will return to this later in this chapter.

I would like to address another element of feasibility. In my view, every discourse rule has its own “feasibility distribution”. That is, every rule has a certain feasibility minimum and feasibility maximum. The feasibility minimum tells us what the lowest possible level of feasibility of the acceptance of (or obedience to) that rule is. This means that it shows us a situation of total infeasibility of a discourse rule. This is a situation in which no single participant accepts (or obeys) the discourse rule in question. Correspondingly, the feasibility maximum tells us what the highest possible level of feasibility of the acceptance of (or obedience to) that rule is. It presents us with a situation in which the maximum level of feasibility is attained: a situation in which the discourse rule is accepted (or obeyed) by every participant in the discourse.

Furthermore, the feasibility distribution of every discourse rule has its own scale: the value difference between the feasibility minimum and feasibility maximum differs per discourse rule. The number of steps that have to be taken to move from the feasibility minimum to the feasibility maximum differ per rule. Moreover, the magnitude of these steps is different for each discourse rule. It is also noteworthy to point out that the feasibility of the rules themselves indirectly determines the feasibility of the preferred outcome. Perhaps we should best see the discourse rules as mutually non-exclusive and mutually non-exhaustive pathways to the preferred outcome. That is, if one rule becomes more feasible, this does not necessarily result in a decrease in feasibility of another rule. Moreover, if one rule is feasible, this does not necessarily imply that all other rules are infeasible\textsuperscript{28}.

\textsuperscript{28} Compare the notions of “probabilistically necessary” and “probabilistically sufficient” causes in a so-called INUS causal scheme in empirical political science. There are no discourse rules that are strictly necessary for the (approximation of the) preferred outcome to become feasible, and multiple different combinations of discourse rules are jointly sufficient for increasing the feasibility of the (approximation of the) preferred outcome. See Mahoney, 2008, pp. 425-427.
Note that there is a broad range of situations that lie in between the feasibility minimum and feasibility maximum. These situations are characteristic of the relative element of feasibility. Whereas the minimum and maximum concern the boundaries of feasibility (full impossibility and full possibility) and thus relate to hard constraints, the points in between relate to soft constraints and therefore concern the degree of feasibility.

Let me accompany these fairly abstract considerations with a more practical example. Take discourse rule (2.1). Approximating reasonable agreement becomes more probable when people become more sincere, because the quality of the discourse is increased in this way. Including more people who are affected by the results of the discourse and have the competence to speak and act results in an increase of discursive quality as well. Also, we will have to determine a feasibility minimum (or point of full infeasibility) and feasibility maximum (or point of full feasibility) for the obedience to each rule first. For rule (2.1), the feasibility minimum is a point whereat no single participant accepts (or obeys) the discourse rule. This means that no single participant is sincere in voicing his needs and concerns. The feasibility maximum for this rule is, correspondingly, a situation wherein every participant expresses his needs and concerns sincerely. An example of a point in between is a situation in which 80 per cent of the participants in the discourse are sincere most of the time. Note that not all participants are sincere. Moreover, the 80 per cent that were mentioned are not sincere all the time, but merely often.

Of course, it is difficult to measure sincerity and to determine at what point between the feasibility minimum and maximum a situation actually is. I suggest that we judge the location of this point comparatively, in an ordinal sense: it is not so much about where the point exactly is, but, rather, about whether the point is nearer to the feasibility minimum (or maximum) than another point. This way, no specific measurement is needed.

The distinction between hard and soft feasibility constraints (or between probabilistic and deterministic elements of feasibility) has already been formulated by other authors (Gilabert and Lawford-Smith, 2012; also see Lawford-Smith, 2013). What is new to the feasibility framework developed here, however, is that multiple kinds of feasibility considerations can be brought together in a larger feasibility framework. This goes beyond the soft/hard distinction, because the framework contains feasibility considerations concerning parts of a theory themselves (here, the discourse rules); feasibility considerations concerning the preferred outcome of the theory as a whole; and the influence the separate parts have on the (path to the) outcome. Moreover, these feasibility considerations can be ranked, because we have these different levels of feasibility on which variation can possibly occur.

I investigated the feasibility of Habermasian discourse ethics above. However, these considerations are not comparable in the sense that we cannot say that economic feasibility constraints, for instance, are more easily removed than cultural ones. With the conception of
feasibility developed above, we can only judge a theory or proposition on a number of aspects, namely the hard and soft constraints I have formulated. It is possible to compare these constraints in the sense that we could say, for example, that economic constraints pose more pressing difficulties for Habermasian discourse ethics than cultural ones. But this does not necessarily mean that mending the economic state of affairs should necessarily have priority over mending the cultural one. The framework developed here can account for these priority considerations, because it allows us to compare multiple different feasibility distributions. It can tell us, for instance, whether investing a lot of money in the economy would lead to a more feasible outcome than doing this in the cultural sector. It provides us with a way in which we can judge whether it is easier or better to prefer certain pathways to the preferred outcome over others. Hence, it is possible to prioritize certain different aspects of feasibility. The feasibility framework I develop here is therefore not only aimed at identifying the constraints on a (path to a) preferred outcome, but also at prioritizing the ways in which these feasibility problems can be resolved. It is a new perspective on feasibility problems and the solutions to them. I will now apply this framework to the Habermasian discourse rules.

5.2 A new set of discourse rules

The feasibility of a pathway to an outcome should be seen as containing two elements: one vis-à-vis the (path to the) preferred outcome, the other internal, as a pathway’s own feasibility distribution. The first concerns impact, the second concerns feasibility. The impact of a discourse rule tells us how important it is for the (path to the) preferred outcome. The feasibility distribution of a discourse rule shows us how difficult it is to increase the feasibility of a discourse rule, and thence the feasibility of the (path to the) preferred outcome. Note that this feasibility distribution contains two feasibility elements, which I also addressed above: on the one hand, the feasibility of the rules themselves, which indirectly influences the feasibility of the preferred outcome; on the other hand, the feasibility of the preferred outcome as such, in a general way.

Let me give an example of how the notions of impact and feasibility should be understood. Recall discourse rule (1.1), which states that “no speaker may contradict himself”. The impact of this rule is the influence that it has on the preferred outcome. This means that we should look at the effect this rule has on the outcome. In other words: how much variation in the preferred outcome does this rule cause? Impact tells us something about to what extent a discourse rule contributes to the preferred outcome of the deliberative democratic project. It can be argued that this rule is important for the outcome, because we cannot attain a certain level of understanding or consensus on the validity of moral norms if we do not use “intrinsically
cogent” arguments (Habermas, 1990, p. 87). Of course, Habermas would argue that all discourse rules are important, because they are necessary presuppositions for a discourse. However, as I argued above, a full reasonable consensus cannot be reached and we should therefore think of the rules in terms of degrees.

Feasibility tells us whether it is possible to attain the preferred outcome; and, if so, to what extent (probability). This does not say anything about the importance of the rule for the preferred outcome. If accepting (or obeying) a discourse rule is feasible, this means that it can be used in practice. In the case of rule (1.1), this would mean that it would be probable (or at least possible) that participants in a discourse do not contradict themselves. If this is the case, then the preferred outcome automatically becomes more feasible as well, because if the chance that people do not contradict themselves increases, the chance of approximating a reasonable consensus also increases; reaching a consensus becomes easier if people do not contradict themselves. What is also important here is that the feasibility of some rules can be enhanced more easily than the feasibility of other rules. Suppose, for instance, that it is easier to increase the probability that participants use the same expressions with the same meanings (rule (1.3)) than it is to increase the probability that participants do not contradict themselves (rule (1.1)). In this case, we can say that the enhancement potential of the feasibility of these rules is higher for rule (1.3) than for rule (1.1).

In order to apply this new framework to the discourse rules, three questions must be answered: what is the impact of each rule? What is the feasibility of each rule (subdivided into current feasibility of the rules on the one hand (hard vs. soft constraints) and enhancement potential on the other)? And how can we rank the rules on the basis of these considerations? These questions will be answered in turn.

5.2.1 The impact of the discourse rules
First, let me restate the discourse rules that Habermas has formulated:

(1.1) No speaker may contradict himself
(1.2) Every speaker who applies predicate $F$ to object $A$ must be prepared to apply $F$ to all other objects resembling $A$ in all relevant aspects
(1.3) Different speakers may not use the same expression with different meaning
(2.1) Every speaker may assert only what he really believes
(2.2) A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so
(3.1) Every subject with the competence to speak and act is allowed to take part in a discourse
(3.2) a. Everyone is allowed to question any assertion whatever
b. Everyone is allowed to introduce any assertion whatever into the discourse
c. Everyone is allowed to express his attitudes, desires, and needs

(3.3) No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in (3.1) and (3.2)

I believe that the impact of the discourse rules consists of three important elements. In my opinion, (the approximation of) a reasonable consensus consists of three main parts: inclusion, the absence of power differences and commitment to the democratic project\(^\text{29}\). These parts will be used together to judge the impact of the rules. For each part, I will assign a value to each discourse rule: a ‘+’ if the discourse rule has a relatively positive influence on a part, a ‘-’ if this influence is relatively negative and a ‘/’ if it does not have a clear one-sided positive or negative influence on the part. Note that it is sometimes argued that these three parts are connected: commitment is influenced by power structures and inclusion, because people who feel discriminated against, left out or disadvantaged will probably not support the project of a society or a discourse community anymore. Power structures and inclusion affect each other, because the more dominant a majority in society becomes, the less it has to mind the interests of the minorities. The majority will exclude minorities from decision-making processes or deliberation more easily. Conversely, excluding minorities leads to larger power differences between the majority and the minorities, because the minorities are unable to “defend” themselves in a legal or political way. Finally, commitment influences inclusion and, indirectly, power structures, because participants who lose the incentive to participate will leave the discourse or become passive in the discourse, which effectively excludes them. Moreover, this leaves the power in the discourse to the participants who carry on with deliberation and can therefore establish the validity of moral norms themselves, without the people who have left the discourse.

There are several rules that have a negative influence on the inclusionary aspect of Habermasian discourse ethics. For instance, rule (2.1) implies that only assertions count in the discourse. People’s assertions tell us what they think, feel or value, but they are not the only way to do so. For instance, body language can tell us whether a person is angry, happy, sad, frustrated etc. Moreover, a sound we make can express emotions or attitudes, but the term assertion does not capture this. This way, it excludes emotions, rhetoric and other important forms of human communication, as we saw earlier. Moreover, it unjustly favours participants who are skilled at deliberating and arguing. Similarly, rule (2.2) is also aimed at reason and excludes all other forms of communication. Some people find it difficult to provide specific reasons for their rejection of certain norms or propositions. They will be excluded from the deliberative project through rule (2.2) if they try to dispute a norm or proposition not under

\(^{29}\)These three aspects are partially based on the work of Smith & Wales (2000) and Rummens (2012).
discussion in their own way of expression. Precisely this kind of discrimination can also be seen in rule (3.1). Deliberation is more than merely speaking and acting. Some people can speak but cannot act in the “normal” sense (e.g. mildly mentally or physically challenged people). Some people can act but cannot speak in the “normal” sense (e.g. mentally but not physically challenged people or people with a speaking disorder\textsuperscript{30}). Of course, there are some people with whom real deliberation would be nearly impossible (e.g. profoundly mentally challenged people or people who do not speak one’s language and also use a kind of body language that is very different from one’s own). However, I contend Habermas has drawn the line too far. The inclusionary aspect of the preferred outcome suffers from this. I believe it is clear that precisely these problems also hold for rules (3.2a) and (3.2b).

Rules (3.2c) and (3.3) do not suffer from these problems and actually promote inclusion. (3.2c) states that everybody may express attitudes etc. An expression is much broader than an assertion or reason. Furthermore, (3.3) promotes inclusion in the sense that everybody is included as a full participant. Both rules aim at ensuring impartiality and do not set any (relevant) boundaries to which people may deliberate and which not\textsuperscript{31}. Therefore, they have a positive influence on the inclusionary aspect of the preferred outcome. Arguably, rules (1.2) and (1.3) also have a positive effect: if participants must use every expression with the same meaning and apply the same predicates to similar objects, this implies that they must agree on the expressions and predicates that are used in the discourse. If everybody must use expressions and predicates in the same way, all (relevant) interpretations of these expressions and predicates that the participants hold must be discussed. For else, agreement on predicates or meanings cannot be reached. It forces people to include all (relevant) interpretations. Finally, I reckon that rule (1.1) has no strong influence on the inclusionary element at all. Self-contradiction does not seem to have anything to do with inclusion.

Now consider the absence of power differences. If we look at the first rule (1.1), the fact that some people are more talented or trained at speaking causes problems, since this rule favours people who have speaking abilities that are above average. People who have these abilities will contradict themselves less often than people who do not have them (or have them to a lesser extent). Rules (2.1) and (2.2) are problematic in the same way, but they also have a positive influence on power structures: (2.1) guarantees that participants cannot use lies, deceit or strategically formulated assertions in the discourse. Rule (2.2) prevents participants from setting the agenda – which is, in my opinion, a form of decision-making power – without a

\textsuperscript{30} Ironcally, Habermas himself has a speaking problem (resulting from palatoschisis). Arguably, precisely the fact that Habermas himself is such a well-known and successful philosopher should have shown him that humans communicate in many other ways than only speech, and successfully.

\textsuperscript{31} Of course, rule (3.2c) implies that participants must be able to express themselves. This could be seen as a boundary, but I contend this is not relevant. For an individual who cannot express him-/herself in any way simply cannot deliberate.
reason. As for rule (3.1), there are, once again, problems of exclusion: power will remain in the hands of those who have the ability to speak and act, and all others will be excluded from the discourse and therefore cannot have any influence on the decisions made in the discourse. Body language, rhetoric, emotions and other forms of expression that are not captured by *assertions* are not considered legitimate ways of expression, which is wrong. The same problem holds for rules (3.2a) and (3.2b), because these are also aimed at *assertions* instead of *expressions* in the broader sense. However, (3.2a) and (3.2b) also have a positive effect on power relations: they give every participant the liberty to say whatever they wish to, and question any assertion they wish to. Thus, theoretically, this forms an obstacle for eloquent people, because even if they are dominant in the deliberative setting, other participants have the power to question all their assertions. Nevertheless, it is problematic that participants will have to choose to use this power. They will possibly not even think about using the power to question others’ assertions – even if they have a good reason to do so – because they live under circumstances of false consciousness and think that the situation they currently live in is the normal one.

Rules (1.2) and (1.3) have a positive influence on power relations in the sense that they necessarily involve all participants in the deliberative process because of the reasons stated above. Moreover, they are not merely focussed on *speech*, but contain the broader notions of *predicate* and *expression*. This implies that participants with an ability to communicate in a different way than through speaking also have the power to influence the discourse. The same holds for rule (3.2c). Finally, rule (3.3) has a clear positive influence on power structures in the discourse. It literally forbids participants to prevent others from expressing themselves on the basis of coercion, and forbids participants to refrain from expressing their concerns because of internal coercion. It therefore guarantees that people do not choose to use any form of power they have to coerce others to express themselves or refrain from doing so.

The third key element of the preferred outcome is commitment. I contend that almost all effects of the discourse rules on inclusion and power structures strongly affect commitment. For example, if some participants have more power than others in the discourse, the disadvantaged people will not have an incentive to carry on and will feel excluded and discriminated against. In turn, if a rule prescribes that as many different people with different abilities as possible be included, or that power differences amongst participants be as small as possible, this will have a positive effect on participants’ commitment to the deliberative democratic project. On the basis

---

32 One could argue that rule (3.3) ensures that eloquent participants may not use their superior speaking abilities to influence the ways in which others express their concerns and needs. However, this would be self-defeating for the discourse, because, first, people cannot simply refrain from communicating in the way they do with their natural or social abilities; and, secondly, every single way of expression would favour the participants who are good at expressing themselves in this very way. This would mean that participants would not be allowed to communicate in *any* way, because every form of communication would favour a certain group of participants. Thus, communication (and deliberation) is impossible.
of these considerations, many of the discourse rules (i.e. (1.2), (1.3), (3.1), (3.2c) and (3.3)) have the same impact on commitment as on inclusion or power structures. As for the rules that have been given different "values" for their influence on inclusion as opposed to power structures, I believe a positive feedback loop will occur. Take rule (1.1), for instance. I have argued that this does not strongly affect inclusion, but that it does affect power structures negatively. If it affects power structures negatively, some people (minorities) will feel discriminated against and will lose their commitment. If people are not committed anymore, they will either leave the discourse or refrain from actively participating in the discourse. This implies that, effectively, less people, and predominantly people from the majority, are included in the discourse. In this way, less commitment probably leads to less inclusion. If more and more people are excluded, the power differences will probably sharpen, which, in turn, will have a negative influence on participants’ commitment. A positive feedback loop becomes apparent. This feedback loop probably also holds for rules (2.1), (2.2), (3.2a) and (3.2b), because the same kind of mechanism can be seen here.

These considerations concerning the impact of the discourse rules are visually summarized in Table 1 below. As can be seen, roughly three groups of discourse rules can be distinguished: first, discourse rules that have a profound negative impact on the preferred outcome (-); secondly, discourse rules that have a mild negative impact on the preferred outcome (-/); thirdly, rules that have a profound positive impact on the preferred outcome (+). As can be seen, most of the discourse rules have a mildly or profoundly negative impact on the preferred outcome. This means that, regardless of the feasibility of either the rules or the outcome, many of the separate rules have a negative effect. We could say that they make the consensus that should follow from discourse less reasonable. In turn, some rules have a positive impact, which increases the reasonableness of the preferred outcome.

Table 1: impact values for separate discourse rules

<table>
<thead>
<tr>
<th>Discourse rule</th>
<th>Influence on inclusion</th>
<th>Influence on power structures</th>
<th>Influence on commitment</th>
<th>Average impact value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.1)</td>
<td>/</td>
<td>-</td>
<td>-</td>
<td>-/</td>
</tr>
<tr>
<td>(1.2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(1.3)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(2.1)</td>
<td>-</td>
<td>/</td>
<td>-</td>
<td>-/</td>
</tr>
<tr>
<td>(2.2)</td>
<td>-</td>
<td>/</td>
<td>-</td>
<td>-/</td>
</tr>
<tr>
<td>(3.1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(3.2a)</td>
<td>-</td>
<td>/</td>
<td>-</td>
<td>-/</td>
</tr>
<tr>
<td>(3.2b)</td>
<td>-</td>
<td>/</td>
<td>-</td>
<td>-/</td>
</tr>
<tr>
<td>(3.2c)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(3.3)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
5.2.2 The feasibility of the discourse rules

The feasibility of Habermasian discourse rules was already discussed in the previous chapter. The general conclusion was that, although most of the discourse rules constitute valuable ideals, these ideals will never be reached or approximated, since the discourse rules are not formulated clearly enough or since they are too demanding for human beings. This analysis, however, only aimed at investigating what feasibility problems exist regarding the discourse rules. Now, I will discuss their potential to enhance their feasibility. This strongly relates to the feasibility distributions of the discourse rules specifically, because these distributions tell us how easy (or difficult) it is to enhance a discourse rule's feasibility. Discourse rules whose feasibility maximum can be approximated more easily should be prioritized over rules whose maximum can only be achieved in a relatively difficult way, because the former provide us with a “quicker route” to an approximation of the preferred outcome than the latter. In this sense, technically speaking, I will now be concerned with the feasibility of enhancing the feasibility of the discourse rules. I will value the rules in the same way as in the discussion concerning their impact.

In my opinion, the first three discourse rules have quite a realistic potential to be made more feasible. In present societies, many people seem to behave in ways that are far from the deliberative ideal (Barabas, 2004; Delli Carpini et al., 2004). However, the ideal that the three rules on the logical level prescribe can be approximated relatively easily: participants can agree on the specific meaning of expressions and objects through deliberation. Also, they can check if these meanings are coherently used and if participants use the same predicates for objects that are similar in relevant aspects. Thus, they can design standards for themselves and create a base of meanings and predicates that everybody can see and tries to stick to. Different studies show that these abilities are more strongly developed through deliberation (Fearon, 1998; Smith & Wales, 2000, pp. 53-54; Barabas, 2004). Moreover, participants will be motivated to develop these standards, because communication will then become easier and decisions will be made more easily. Furthermore, on first glance, it might seem that rules (1.2) and (1.3) already prescribe that participants’ expressions be coherent and consistent, which suffices for the absence of self-contradiction (1.1). However, Habermas is concerned with “intrinsically cogent arguments with which we can redeem or repudiate claims to validity.” (Habermas, 1990, p. 87). This goes beyond the consistency with which we use the meanings and predicates that we have agreed upon; it also concerns the consistency of an argument itself. Even if an argument is made with meanings and predicates that all agree upon, it can be inconsistent, and this is precisely what Habermas means.

For rule (2.1), enhancing feasibility seems well possible, too. As I argued in the previous chapter, some participants probably cannot know what they really believe, or even do not want
to know. However, several authors have argued that this ability can be enhanced precisely through deliberation, as with the rules on the logical level. Participants will develop a stronger sense of mutual respect (Gutmann & Thompson, 1990; Warren, 1996), a more open view to others’ standpoints and positions (Manin, 1970; Barabas, 2004) and a more active notion of citizenship, which shows them that values and preferences can, and sometimes should, be changed (Baynes, 1995; Young, 1996). So although participants begin deliberation under circumstances of “false consciousness”, this could lessen if participants deliberate longer and/or more often. The same goes for rule (2.2), I assert: deliberative “training” results in people not only being able to accept (or obey) the rule, but also wanting to follow it for the sake of the discourse.

Let us turn to the rules on the process level. I have argued that rule (3.1) is too vague to draw a (justifiable) boundary between participants and non-participants. Although it should be altered, I discuss the rule Habermas has formulated himself here; I will discuss altering the rule later. There is an enhancement potential for this rule, but it is small and possibly discriminatory. We could try to educate people in the art of (public) speaking and arguing, so that they will be better participants in a discourse. However, education is always, at least partially, culturally biased, and many participants will not realize this. Moreover, some people will learn faster or better than others, which creates further ability inequalities and has a negative influence on power structures, inclusion and commitment. Therefore, I believe rule (3.1) cannot be justifiably enhanced in the form it is formulated. In my short analysis of rules (3.2a) through (3.2c), I argued that the basic thought of these rules is right and that they seem to be quite feasible. Judging the rules as they are stated by Habermas, I believe a standard for the relevance of expressions in the discourse will gradually come into existence: acceptance of (or obedience to) rules (1.1) through (1.3) will lead to agreement on a set of expressions, meaning and predicates that are used publicly. If participants talk about the language they use and determine what language and expressions are permitted, that implies that they create such a standard for acceptable versus non-acceptable language and expressions. Thus, deliberation “automatically” enhances the feasibility of the rules. However, it is a pity that rules (3.2a) and (3.2b) do not account for this enhancement potential adequately enough: they only concern assertions, not expressions. This is what should be altered concerning these discourse rules. Finally, I concluded rule (3.3) is infeasible because of the workings of power structures, group loyalty, differences in talent for speaking, cognitive dissonance reduction and, importantly, because it does not discern multiple kinds of coercion or inequality. I assert this rule must also be altered. Judging it from the current perspective, I believe its feasibility cannot be enhanced in a strong sense. Although people become more “civil” and mutually respectful through discourse (Gutmann & Thompson, 1990; Baynes, 1995; Warren, 1996; Young, 1996), I believe that some forms of power structures
or elements of human communications (e.g. natural talents, luck or bad luck or reputation) are ineradicable. Moreover, differences (virtually necessarily implying power differences) will always exist, because we need them to discern and to make choices (cf. Mouffe, 2000; 2005). Therefore, the rule has some potential, but only to a very limited extent. It is too idealistic to even approach the feasibility maximum, I contend.

On the basis of these considerations, Table 2 can be created. As can be seen, roughly four groups of discourse rules can be formed on the basis of the average feasibility value: rules with a profound negative effect (-), a mildly negative effect (-/), a profound positive effect (+) or an effect that is neither clearly positive nor negative (/). The rules that have been given positive values are rules whose feasibility and/or potential of feasibility enhancement is high. Correspondingly, rules with negative values have a low level of feasibility and/or potential of feasibility enhancement. The outcomes of this analysis say something about the feasibility of each discourse rule specifically.

Table 2: feasibility values for separate discourse rules

<table>
<thead>
<tr>
<th>Discourse rule</th>
<th>Present feasibility of rule</th>
<th>Potential of feasibility enhancement</th>
<th>Average feasibility value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.1)</td>
<td>-</td>
<td>+</td>
<td>/</td>
</tr>
<tr>
<td>(1.2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(1.3)</td>
<td>-</td>
<td>+</td>
<td>/</td>
</tr>
<tr>
<td>(2.1)</td>
<td>-</td>
<td>+</td>
<td>/</td>
</tr>
<tr>
<td>(2.2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(3.1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(3.2a)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(3.2b)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(3.2c)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(3.3)</td>
<td>-</td>
<td>/</td>
<td>-/</td>
</tr>
</tbody>
</table>

5.2.3 Ranking the discourse rules

In order to rank the rules based on their importance concerning feasibility considerations, we must combine the feasibility and impact values. Impact and feasibility are related: if a rule has a large impact on the preferred outcome, it makes the outcome more feasible. If a rule is feasible (or has the potential to become feasible) compared to the other rules, this also, indirectly, increases the feasibility of the (path to the) preferred outcome. Figure 1 below combines the feasibility and impact values in a dot plot. The upper left quadrant contains rules that have a large impact but are relatively infeasible. The upper right quadrant contains rules that have a large impact as well as a high feasibility value. Correspondingly, the lower left quadrant presents
rules that are neither feasible nor influential and the lower right quadrant contains rules that are relatively feasible but lack a strong impact on the preferred outcome.

How can we group these rules into a ranking? Naturally, rules that are both feasible and have a large impact are the most important in terms of the feasibility of discourse ethics as a whole. Correspondingly, rules that are neither feasible nor influential are the least important. As for the other rules, the further we move from the upper right to the lower left quadrant (be that either through the lower right or the upper left quadrant), the less important the rules become. On the basis of this interpretation, the discourse rules can be ranked. These rankings are summarized in Table 3 below. Rank 1 is the highest rank, rank 5 the lowest.

The way I have ranked the discourse rules using a broader feasibility framework can be applied to other aspects of Habermasian discourse ethics or even other theories as well. As I have said, I applied this framework to the discourse rules specifically because these rules are a relatively clear and structured part of Habermas’s theory. There is no room in this thesis for applying this framework to other parts of this thesis. The main point here is to show that thinking in terms of feasibility in the way I proposed above provides us with a more complete set of feasibility considerations that is still applicable to a wide number of other (parts of) theories. Ranking the discourse rules in this way shows us that if we wish to enhance the potential of discourse ethics as a whole, or, in other words, increase the feasibility of (the path to) its preferred outcome, we should best start with applying, implementing or striving after the rules that have been given a high rank – rules (1.2), (1.3) and (3.2c), for instance.

Table 3: a ranking of the discourse rules

<table>
<thead>
<tr>
<th>Rank</th>
<th>Discourse rule</th>
<th>Impact average</th>
<th>Feasibility average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1.2)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>(3.2c)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>(1.3)</td>
<td>+</td>
<td>/</td>
</tr>
<tr>
<td>3</td>
<td>(2.2)</td>
<td>-/</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>(3.2a)</td>
<td>-/</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>(3.2b)</td>
<td>-/</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>(3.3)</td>
<td>+</td>
<td>-/</td>
</tr>
<tr>
<td>4</td>
<td>(1.1)</td>
<td>-/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>(2.1)</td>
<td>-/</td>
<td>/</td>
</tr>
<tr>
<td>5</td>
<td>(3.1)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
5.2.4 Altering the discourse rules

I stated above that I believe some discourse rules should be altered. Here, I wish to briefly propose alternative formulations of the rules that I think should be altered.

Rule (3.1) is too vague to draw a line between participants and non-participants and excludes people that should be included. I believe we should include more people than merely those who have “the competence to speak and act”. “The competence to express themselves in a way that is understandable for the other participants in the discourse” is a superior alternative to the original formulation, I assert. There will still be some problems with this formulation –
that is inevitable. For instance, people with a profound mental or physical handicap may be unable to even express themselves in a way that is understandable. The best solution to this problem seems to be representation. However, as I argued before (chapter 4), it would be best if this representation is not restricted to merely profoundly physically/mentally handicapped people, but extends to (all) other groups of the population “fit” for deliberation as well. I admit that the formulation provided here is still relatively abstract. However, the rule becomes more inclusionary, and will therefore also positively affect power relations and commitment issues.

Rules (3.2a) and (3.2b) contain one small problem: they merely focus on *assertions*. I propose to simply change “assertions” into “expressions”, the latter of which is a broader and more inclusive term.

As for rule (3.3), I believe we must introduce several sub-rules, because some forms of coercion are graver than others. Physical violence and grave mental and psychological abuse are more severe forms of harm than mild mental abuse, offensive behaviour and disgust towards primary objects (cf. Nussbaum, 2010). Alas, I am unable to develop a full conception of degrees of harm here.

The rules I would like to propose, in a preliminary sense, are the ones below. I have written the alterations I have touched upon, as well as some minor other ones, in italics.

1.1 No participant may contradict himself
1.2 Every participant who applies predicate $F$ to object $A$ must be prepared to apply $F$ to all other objects resembling $A$ in all relevant aspects
1.3 Different participants may not use the same expression with different meaning
2.1 Every participant may express only what he really believes
2.2 A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so
3.1 Every subject with the competence to express him-/herself in a way that is understandable for the other participants in a discourse is allowed to take part in a discourse
3.2 a. Everyone is allowed to question any expression whatever
       b. Everyone is allowed to introduce any expression whatever into the discourse
       c. Everyone is allowed to express his attitudes, desires, and needs
3.3 No participant may be prevented by coercion from exercising his rights as laid down in (3.1) and (3.2), whereby
       - Physical abuse and grave mental or psychological abuse have a higher priority than mild mental or psychological abuse
       - Mild mental or psychological abuse has a higher priority than offensive behaviour or disgust towards primary objects
6 CONCLUSION

In this last chapter, a short summary of this thesis will be given and the implications of my research for the question I set out to answer at the beginning will be discussed. A short critical reflection on the completed research will follow. Finally, several suggestions for future research will be presented.

6.1 Summary and implications for research question

Jürgen Habermas's moral theory of discourse ethics is one of the main contemporary deliberative democratic political theories. Although it remains influential and quite successful now, his theory has been met with many diverse points of critique as well. One of these main points of critique states that discourse ethics is too strongly focussed on ideals and/or pays too little attention to feasibility considerations. Precisely because Habermas claims that his theory is a critical theory that is applicable to as well as partially representative of contemporary society, the feasibility of his theory is of great importance. Feasibility considerations, applied to Habermas's theory, constitute the main concern of this thesis. The main elements of Habermas's theory as well as several strands of critique on it were discussed earlier. Moreover, a conception of feasibility was presented, partially based on the work of Gilabert & Lawford Smith (2012) and Lawford-Smith (2013). There can be two kinds of constraints on feasibility, namely hard and soft constraints. The former are deterministic and can render the preferred outcome of a theory impossible; the latter are probabilistic and render the outcome more or less probable. After considering the feasibility of the main elements of discourse ethics in this bipartite way, it was concluded that there are several important feasibility issues for Habermasian discourse ethics. These issues predominantly concern its impossibility in terms of momentum, its logical inconsistencies, the impossible biological and/or psychological capacities that are demanded of individuals, the lack of motivation to start or participate in a discourse and the formulation of the discourse rules specifically. Focussing on the infeasibility of the discourse rules, a new feasibility framework was developed and applied to Habermas's theory. Feasibility, it was argued, should be seen as a part of a broader framework that contains two main elements: impact and feasibility. Impact affects the feasibility of the preferred outcome and the feasibility of the discourse rules affects both the rules themselves as well as the (path to the) preferred outcome. Impact, in the case of Habermasian discourse rules, was defined as the importance of the rules for the three main parts of the (path to the) preferred outcome, viz. inclusion, power structures and commitment. Feasibility ought to be seen as a bipartite concept: on the one hand,
it is concerned with what is *problematic*, be it now or in the (near) future. On the other hand, it relates to the *enhancement potential* of parts of a theory. In other words, we should look at both the problems a theory has as well as how probable it is that these problems can be solved. The application of this new framework led me to infer that a possible way in which the feasibility of Habermas’s theory can be enhanced is to rank the discourse rules. Five groups of discourse rules were created, based on their values for impact as well as feasibility. Moreover, it was argued that several alterations in the formulation of the discourse rules could have a positive effect on feasibility issues as well.

At the outset of this thesis, a research question was formulated: "*How can Habermasian discourse ethics be made more feasible?*** I believe that applying the feasibility framework developed here to Habermasian discourse ethics (in this case, the discourse rules specifically) shows us that the notion of feasibility is broader than it is often presented in the literature. Furthermore, it lets us think beyond the boundaries of *problematique*, since it also stimulates us to consider the *resolution/enhancement potential* of a (part of a) theory. This fairly abstract answer to the research question is accompanied by a more practical example of the application of the framework, namely the ranking of the discourse rules. The feasibility of Habermasian discourse ethics can be increased by utilizing a useful and complete conception of feasibility, based on impact and feasibility, to judge and enhance it, as well as by ranking the discourse rules as a practical example of such an enhancement. Applying a more complete conception of feasibility to Habermas’s theory sheds light on aspects of feasibility that have been neglected in previous literature – impact being one of the most important ones. This way, investigating the feasibility of Habermas’s theory (and other theories) become more precise and more valid. Furthermore, by ranking several parts of a theory on the basis of feasibility and impact considerations, it can be judged which theory parts and corresponding decisions or implementation procedures should be prioritized over others. This makes effectively applying and implementing political theories easier and more feasible. Therefore, the framework developed here is not only a new and more complete way to consider the feasibility of a theory, but also a way to indirectly *increase* the feasibility of a theory (or parts of it).

### 6.2 Critical reflection

This thesis has several weaknesses. First, as is the case in many articles or theses, not all topics that are in some way relevant to the research question could be discussed. Scarcity of time and the lack of room for more words negatively affects the extensiveness of the topics that are discussed here. This is the reason why it is impossible to apply the new feasibility framework to all main elements of Habermas’s theory in this thesis. Secondly, I have not used any primary
empirical data to support my arguments, but, rather, only secondary sources or philosophical arguments. Thirdly, Habermasian discourse ethics is the only theory that is discussed here. I believe the feasibility framework I developed earlier will be a useful tool for scholars considering the feasibility of other theories, but alas, I do not have the opportunity to do suchlike research here. Finally, the ways in which the feasibility of Habermasian discourse ethics – and its rules – can be enhanced have not been tested empirically. My argument is predominantly theoretical and it is unsure whether the suggestions I proposed are useful or applicable in an empirical sense.

Nevertheless, I believe I have reached the goal I formulated at the beginning. My main concern was to alter the way in which scholars think about and apply conceptions of feasibility. The change in thinking I opt to effect is largely theoretical, because in my opinion, feasibility considerations start with conceptualizing feasibility. I have done precisely this, as well as given an example of how the new framework developed here could be applied to theories. Therefore, I believe the problems regarding the lack and/or quality of the empirical work in this thesis are not as grave as they seem.

6.3 Suggestions for future research
First and foremost, it would be interesting to apply the framework that was developed earlier to more (and other) theories in order to see whether it is not only useful in a theoretical sense, but also in an empirical one. Moreover, this would provide a more reliable way in which we can judge the usefulness of the framework. Secondly, it would be important to actually apply this framework to Habermasian (or other deliberative democratic) theories in reality. This way, the feasibility and validity of the framework itself can be assessed. Thirdly, scholars could apply the framework to other elements of Habermasian discourse ethics. The discourse rules that the framework was applied to here form a relatively clear and structured part of Habermas's theory, I contend. If the new feasibility framework is used to explore the feasibility of other main elements of his work as well, a more complete conclusion concerning the feasibility of Habermas's theory can be drawn.
REFERENCES


Jürgen Habermas’ discours ethiek is een bekende deliberatief-democratische theorie die vaak geloofd wordt vanwege haar idealen. Verscheidene auteurs hebben echter beargumenteerd dat wat deze theorie voorschrijft en vooronderstelt niet haalbaar (feasible) is, of in ieder geval niet haalbaar genoeg. Een theorie kan weliswaar waardevolle idealen bevatten, maar als haar doelen en assumpties onhaalbaar zijn, is zij ongeschikt voor de praktijk. Het doel van deze scriptie is het onderzoeken van een manier waarop Habermas’ theorie haalbaarder gemaakt kan worden. Na het constateren van de zwakke punten in de theorie, zowel in algemene zin als specifiek betreffende haalbaarheidsproblemen, wordt er beargumenteerd dat haar haalbaarheid moet toenemen. Een nieuwe conceptie van het begrip ‘haalbaarheid’ wordt ontwikkeld, gebaseerd op bestaande literatuur over haalbaarheid. Het gebrek aan haalbaarheid in Habermas’ theorie kan met de toepassing van deze conceptie worden aangetoond. Daarna wordt beargumenteerd dat zelfs de ontwikkelde conceptie van haalbaarheid niet geheel accuraat is, omdat haalbaarheidsproblematiek meer omvat dan slechts de elementen die in de bestaande literatuur worden besproken. Omwille van deze reden wordt er een nieuw haalbaarheidskader ontwikkeld, waarin zowel de haalbaarheid van afzonderlijke delen van een theorie als de geprefereerde uitkomst van die theorie een plaats hebben. Dit nieuwe kader wordt vervolgens toegepast op een belangrijk deel van de discours ethiek: de discoursregels. Het nieuwe haalbaarheidskader biedt de mogelijkheid om een theorie, of enkel verschillende harer delen, op een completere manier op haar haalbaarheid te beoordelen dan tot nu toe in de wetenschap gepoogd is. Het kader kan niet alleen toegepast worden om de haalbaarheid van Habermas’ theorie te beoordelen en te verbeteren, maar kan op dezelfde manier op andere (politieke) theorieën toegepast worden.