Justification and Moral Cognitivism

An Analysis of Jürgen Habermas's Metaethics

Author: Johan Elfström
Supervisor: Professor Elena Namli

Abstract

In this thesis, I scrutinise and interpret Jürgen Habermas's claim that justification of moral norms necessitates cognitivism. I do this by analysing the general idea behind his discourse theory of morality and then his metaethics. From there, I examine the non-cognitivist theory called prescriptivism as set out by Richard Hare to see if his account of moral reasoning is able to counter Habermas's claims and thereafter, I examine some criticism against his concept of communicative action. I also engage with the discussion on how to define cognitivism: that is, whether the line should be drawn between moral realism on the cognitivist side, and constructivism on the other, or if cognitivism can include constructivist theories too. I propose that it should, provided that it allows moral statements to be truth-apt and express a mental state like that of belief. Following this definition, I argue that Habermas can be labelled a cognitivist and finally, I conclude that Habermas argument does not hold under scrutiny.

Table of Contents

Johan Elfström

1. Introduction

Considering the philosophical dimensions of human rights theory, their justification is one of the more difficult problems. Claiming universality, human rights need either be compatible with many, sometimes rather distinctly different moral traditions, or shown to be superior to them. Attempts of justification often include aspects of both strategies; showing that, and how, human rights can be made compatible with a certain religious tradition for instance, as well as trying to show that it is, say, more rational of someone to accept the claims by the human rights approach than the moral beliefs currently held. Both strategies come with their own particular issues however. In the first case, one might have to specify the rights in such a way that they are rendered insignificant in either scope or impact, or both; in the latter case, the project may manifest itself in terms of force rather than rational argument.[1] This latter issue is, broadly speaking, the issue of this thesis.

One family of philosophies claiming that justification can be provided through deliberation is commonly referred to as *discourse ethics*, within the development which the German philosopher and sociologist Jürgen Habermas has been a key figure. One of the key issues which divides theorists within this tradition is that of discourse as argumentation with the purpose of reaching a rationally motivated consensus. While Habermas has maintained the importance of reason, others have been sceptical about the plausibility of such a project.[2]

The problem of justification is in many respects one of normative moral philosophy, and as the title of this thesis suggests, my concern is with that of metaethics, not focusing so much on what is good or right and why, but *what it means for it to be so*. This metaethical question has several dimensions, all begging different kinds of answers: one might ask whether it is possible to have knowledge of moral matters, or the related question whether moral propositions can be true; furthermore, one might focus on the meaning of the words of which the moral vocabulary consists and what their function is; or one might wonder whether values, like facts, has an existence which is mind-independent. These questions correspond to three problems that philosophers interested in this area of study concern themselves with: the epistemological

---

[1] Seyla Benhabib, p. 202; Jeffrey Stout, *Democracy and Tradition* (Princeton: Princeton University Press, 2004)., p. 76. Benhabib criticises John Rawls for holding a too minimalist view of human rights, Stout criticise him for trying to impose a western liberal perspective on other traditions. This is a common criticism against liberalism, and Rawls is not its only subject.

[2] Benhabib has rejected the search for consensus; as did Iris Marion Young, who wrote on discourse ethics from the viewpoint of deliberative democracy and whom did not have much of a positive outlook on the prospects of rationally motivated agreements; Jenny Ehnberg too has criticised the plausibility of such epistemological universalism. See Seyla Benhabib, *Situating the Self*, p. 37; Jenny Ehnberg, *Globalization, Justice, and Communication* (Uppsala: Acta Universitatis Upsaliensis, 2015)., p. 164, 259; Iris Marion Young, *Justice and the Politics of Difference* (Princeton: Princeton University Press, 1991)., p. 106f.

problem, the semantic problem, and the ontological problem.[3] All of which will, although to varying degrees, be touched upon in this thesis. Considering the focus of discourse ethics, it is important that the different metaethical levels – or problems – are given an account which motivates, and even better, requires argumentation. In Habermas's view, such an account is a *cognitivist* one, meaning that moral judgements, in some regard, are truth-apt. As will be clear already in the next section, and even more so throughout the thesis, this is a rather blunt explanation of this position, but for the present purpose it will suffice.

For Habermas, it is not enough to claim that cognitivism gives the better answer to these questions, but he argues that cognitivism is, in fact, necessary for one to be able to rationally justify moral norms. Non-cognitivist accounts, he claims, reduce moral statements to experiential or intentional sentences, or imperatives, and thus '... in a single blow, non-cognitivist approaches deprive the sphere of everyday moral intuitions of its significance'.[4] Since these forms of sentences cannot raise claims to any form of validity, especially not truth, there is no need to argue about moral judgements. There is no need to argue since agreement on moral statements is merely agreement on moral principles which cannot themselves be grounded.

1.1. Cognitivism and Rational Justification

As often is the case with philosophical concepts, there are rarely merely one plausible understanding of it, and motivating one's definition of choice is thus an important part. Carl-Henric Grenholm proposes, in his book *Etisk teori*, that cognitivists hold that moral propositions '... ger oss kunskap om sakförhållanden [which translates into: ... gives us knowledge about states of affairs]'.[5] Seyla Benhabib, on the other hand, suggests that cognitivism means that one can justify to others, using good reasons, that some principle is morally right.[6] In an article on moral cognitivism in the Stanford Encyclopedia of Philosophy, Mark van Roojen writes that cognitivism means '... that moral sentences are apt for truth or falsity, and that the state of mind of accepting a moral judgment is typically one of belief'.[7] To my mind, van Roojen's definition is superior since he is highlighting two aspects which cannot be easily left out by a cognitivist

---

[3] Carl-Henric Grenholm, *Etisk teori*, p. 21, 23, 25.
[4] Jürgen Habermas, 'Discoure Ethics: Notes on a Program of Philosophical Justification', in *Moral Consciousness and Communicative Action* (Cambridge: Polity Press, 1990), pp. 43–115., p. 55.
[5] Carl-Henric Grenholm, *Etisk Teori* (Lund: Studentliteratur, 2014)., p. 22
[6] Seyla Benhabib, *Situating the Self: Gender, Community, and Postmodernism in Contemporary Ethics* (Cambridge: Polity Press, 1992)., p. 49f.
[7] Mark van Roojen, 'Moral Cognitivism vs. Non-Cognitivism', in *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2016)
<https://plato.stanford.edu/archives/win2016/entries/moral-cognitivism/>.

moral theory. Further down in his article, he specifies two theses of cognitivism[8] which I take to correspond the above definitions rather well: the first one is that moral judgements have propositional content, and the latter that moral judgements '... fall on the cognitive side of the cognitive/non-cognitive divide'. Considering this, I take the first criterion to touch upon the definition that Grenholm provides, whereas the second, upon that which is suggested by Benhabib.

Drawing a line between cognitivist and non-cognitivist theories are easier said than done, it seems. As van Roojen notes, not all who calls themselves non-cognitivists deny both theses of cognitivism, and as is seen above, those that claim to be cognitivists does not do that either. Furthermore, philosophers that claim to be non-cognitivists might accept both theses although deny that any one of them are the primary function of moral propositions.[9] Some non-cognitivists might also accept truth-aptness if truth is defined in some deflationary sense. As I will understand cognitivism in this thesis, I will hold that a cognitivist must accept both criteria: truth-aptness and its correspondence to a cognitive mental state that is. I will also hold that a cognitivist will take the function of both aspects to be central to moral propositions. I will assume only a minimal definition of truth, similar to the deflationist theory, although I will not claim that the deflationist theory is a reasonable understanding of truth. I do this only to suggest a definition of truth-aptness that does not discriminate against more substantial definitions of truth. My understanding of cognitivism will be elaborated upon during the course of the thesis and will be fully explicated towards its end. The reason for this is that this discussion about the definition of cognitivism presupposes a certain understanding which some readers will not have, and I am unable to give, at this early stage.

I will only give a very broad account of what is meant by justification too. It might be rather simply stated that justifying a claim, one provides reasons for thinking something to be, say, true or right. One can justify a belief to oneself, figuring out the answer to the question what one should believe or do, it can also refer to a process of argumentation consisting of two or more participants trying to convince each other about the correctness of a claim. Indeed, trying to justify a contested claim is not very helpful unless it meets a certain standard. Hence, something needs to be said about how one might distinguish between one which does, and one which does not do that. At the most basic level there must be some form of relation between

---

[8] Mark van Roojen, 'Moral Cognitivism vs. Non-Cognitivism', see section 1.1. 'Two Negative Constitutive Non-Cognitivist Claims'.
[9] Mark van Roojen, 'Moral Cognitivism vs. Non-Cognitivism', see section 1.1. 'Two Negative Constitutive Non-Cognitivist Claims'.

the claim and the reasons given in its favour so that the claim follows, say, deductively or inductively from the given reason.[10] If no such relation can be established between a claim and a reason, that is, if the reason adduced does not support the claim, it does not provide much in terms of a rational justification. I will also assume that in order to be able to rationally justify a claim, a speaker must be entitled to believe that the reason given is true, or in case of a normative reason, that it is right. It does not seem reasonable to say that a speaker has rationally justified a claim of theirs if they are not sincerely believing that either the claim or the reason is true or right.[11]

This distinction between justification and rational justification is thus made in order to distinguish between a justification consisting of reasons that may rationally motivate a hearer, and reasons which does not but instead relies on the speaker's power. But also, when it is of the former sort, in order to distinguish between one in which the speaker sincerely believes the content of and one which is given aiming to deceive the hearer. This means that if a speaker gives reasons which they themselves does not believe in but which the hearer, according to their worldview, might rationally accept, it is not a rational justification. The same holds for instances where the speaker withholds relevant information. It is not impossible to think of such a situation where a speaker is, say, deliberately withholding important reasons or tell lies to make the hearer accept the claim. It does not follow, of course, that one must, in order to give a rational justification, give a complete account of every aspect pertinent to a claim. Perfect information is not something anyone has. If one does know of a reason which might either increase or decrease the likelihood of the claim, then one cannot ignore it.

1.2. Previous Research

Having made clear the overarching problem of this thesis, and having made some first necessary restrictions of its scope, I will narrow it further in this second section by giving some examples of similar research, and thereby drawing the reader's attention to the particular issue with which I am concerned. Habermas has been the subject of quite a few books and articles that deals both with the entire spectrum of his philosophy, as well as more or less specific aspects and problems in it. Some of these books are textbooks. One example is *Habermas: A Very Short Introduction* which is written by James Gordon Finlayson. While trying to explain Habermas's philosophy,

---

[10] As Robert Alexy argues, Stephen Toulmin suggested that a valid moral inference has an entirely different form and is neither deductive nor inductive; as is also quite often stressed, by Benhabib for instance, moral inferences has restrictions in such a sense that one cannot infer an 'ought' from an 'is'. See Robert Alexy, *A Theory of Legal Argumentation* (Oxford: Oxford University Press, 1989)., p. 83; Seyla Benhabib, *Situating the Self*, p. 30.

[11] Tomas Ekenberg and others, 'Berättigande', in *Kunskapsteori: En Introduktion till Vetande, Berättigande Och Sanning* (Lund: Studentlitteratur, 2014), pp. 33–51., p. 43.

clearly and concisely, this book also has a distinct critical approach trying to show some problems in his theories and discuss possible solutions. Another one is *The Philosophy of Habermas* by Andrew Edgar, who also tries to explain Habermas's entire philosophy from his early years until today, this book does not seem to have as much of a critical approach but do discuss some commonly made objections and discuss varying interpretations. The list of books in this genre can certainly be made even longer.

The Linguistic Turn in Hermeneutic Philosophy is a book by Cristina Lafont, whom in its third part analyses Habermas's idea of communicative rationality, trying to solve some problems she has identified in his epistemology by interpreting it from an internal realist perspective. The problem is, Lafont claims, that Habermas's antirealist position regarding the "… connection between meaning and validity",[12] which is the ground for the non-epistemic interpretation of truth-claims, as well as the claims to moral rightness, has two equally intolerable consequences. Either it implies a metaphysical position that is inconsistent with fallibilism, or a consistently antirealist one, in which one must accept relativism. Thus, her aim is to give an interpretation of Habermas epistemology that do not result in having to take up any of these positions.

A similar undertaking can be found in Steven Levine's article 'Truth and Moral Validity: On Habermas' Domesticated Pragmatism'. Here, Levine argues that Habermas, when he abandoned his early epistemology, to escape the threat of contextualism concerning truth, his new position needed him to domesticate' his pragmatism in order to be able to keep his distinction between truth and rightness. Levine is critical of this manoeuvre, since an undomesticated pragmatism would allow rightness to admit to truth. James Scott Johnston claims the contrary. In his article 'Is There a Need for Transcendental Arguments in Discourse Ethics', he urges Habermasians to abandon the discourse theory of morality's universalist claims. Habermas fails, Johnston says, to reconcile the pragmatism of George Herbert Mead with Immanuel Kant's transcendental idealism. While Johnston, like Levine, argues in favour of pragmatism, they take opposing sides regarding Habermas's cognitivist position.

The question of the acceptability of cognitivism is, as explained by Grenholm, a concern for various philosophical fields: philosophical logics being one important example. In 1965, the article 'Assertion', written by Peter Geach, was published in the journal *The Philosophical Review*. Geach was a professor of logic, whom, in this article, claims that non-cognitivism creates a problem for moral argumentation. The problem is that the normative character of

---

[12] Cristina Lafont, *The Linguistic Turn in Hermeneutic Philosophy* (Cambridge: MIT Press, 1999)., p. 280.

words such as 'right', 'wrong', or as in Geach's example, 'bad', cannot be maintained in all contexts in which it is likely to appear. His example is as follows:

1. If doing a thing is bad, getting your little brother to do it is bad.
2. Tormenting the cat is bad.
3. *Ergo*, getting your little brother to torment the cat is bad.[13]

Since the cognitivist can claim that the word is descriptive each time it appears, the problem is for the non-cognitivist, denying that moral sentences have a descriptive function, to show that the meaning of the word "bad" does not change between the first and second premise. As long as they are unable to do that, non-cognitivism is an unacceptable position.

1.3. Purpose and Question

The purpose of this thesis is to interpret and scrutinise the claim made by Habermas regarding the relation between cognitivism and the rational justifiability of moral statements. More specifically, his claim that rational justification of such statements necessitates cognitivism. In order to do this, I ask the following questions:

1. How does Habermas justify his claim that the rational justification of moral statements necessitates cognitivism?
2. How does Habermas's argument for the necessity of cognitivism fare against the objections given by his critics?

1.4. Sources and Method

I will use the essay 'Discourse Ethics: Notes on a Program of Philosophical Justification', published in the book *Moral Consciousness and Communicative Action*, as my point of departure. Originally published in the first half of the 1980's, this essay is an early, albeit not the earliest, presentation of the discourse theory of morality, or discourse ethics, as it was called at the time. I will also frequently refer to the book *The Theory of Communicative Action*, published the year after the previously mentioned book, and is often considered to be Habermas's flagship. Focusing on its first volume, it provides an in-depth account of his understanding of rationality in relation to his theory of communicative action.

There are, of course, also a set of secondary sources. To consider whether Habermas's claim is plausible or not, I must examine a non-cognitivist theory too in order to see whether Habermas's claim holds. For this I have chosen Richard Hare whose theory seems to be the

---

[13] Peter Geach, 'Assertion', *The Philosophical Review*, 77.4 (1965), 449–65., p. 463.

non-cognitivist theory that is most likely to disprove Habermas's claim. I will be focusing on his book *The Language of Morals*, but later works *Freedom and Reason* and *Moral Thinking* has also a role to play. Along with Hare, I will also analyse relevant aspects of Carl-Henric Grenholm's ethical theory, as presented in *Etisk teori*, in which Hare's prescriptivism is modified. The reason for this, is that his modifications takes prescriptivism in a rather interesting direction, closer to that of Habermas's position. Central to this thesis is also Stephen White's book *The Recent Works of Jürgen Habermas*. This book will be featured throughout the thesis. Furthermore, I must mention the books *Situating the Self* by Seyla Benhabib, and *Polis and Praxis* by Fred Dallmayr among this secondary literature. Both has presented some criticism of Habermas's theory relevant to my thesis. Lastly, I must mention the book *Inquiries into Truth and Interpretation* by Donald Davidson, and the article 'Moral Cognitivism', written by Hallvard Lillehammer as they are key components for elaborating upon my understanding of cognitivism.

The answer to my research question will, as said, be provided in three parts: *firstly*, I will explain the discourse theory of morality, focusing on Habermas's cognitivism and its central place for the justification of moral and political principles and norms. Explaining his position, I am not only interested in the function that cognitivism has in his theory, but in theories of justification in general. Thus, the task here is to provide a reasonable interpretation of the discourse theory of morality, meaning that it is logically coherent and, at the same time, remains close to Habermas's intentions. I will treat the first criterion as a general directive which may be trumped by the second one. Put differently, I am working with the *prima facie* assumption that Habermas theory is coherent, and thus, that, if several interpretations are possible, the most coherent one is also the one that Habermas has intended. I will choose the less coherent interpretation if it can be sufficiently argued that it is the one closer to his intentions. Such arguments can consist, among other things, of indications in his own texts, as well as interpretations provided in others' work. *Secondly*, I will consider some objections to Habermas's claims as they have been presented, and my reading of his critics: Richard Hare, Carl-Henric Grenholm, Seyla Benhabib, and Fred Dallmayr, will of course follow these considerations too. Having already indicated some problems in Habermas's theory, I proceed with a more detailed analysis of these problems. *Lastly*, I consider whether the arguments uncovered are able to refute Habermas's claim, and thus answer the research question.

There are, in broad terms, three possible conclusions: first (1), that Habermas, with regards to the limited scope of this thesis, seems to be correct in his claim that cognitivism is necessary for rationally justifying moral norms; secondly (2), that it, contrary to what Habermas

is arguing, is not necessary to accept cognitivism generally, but that the discourse theory of morality relies on it, and, thirdly (3), that it is not necessary to accept cognitivism in either case. I will deem (1) to be the case, if, given the evidence given in this thesis, unless cognitivism is true, normative reasons cannot rationally motivate anyone to accept the norm or statement to which it refers as valid. If (3) is to be the case, considering the reasons given in this thesis, normative reasons seems able to rationally motivate someone to accept the norm to which a moral statement refers as valid regardless of the philosopher's stance on cognitivism. For (2) to be the case then, given the evidence presented in this thesis, cognitivism is necessary for the discourse theory of morality, although this conclusion cannot be generalised.

It should be noted that while conclusion (1) implies the rejection of non-cognitivism, (3) and (2) are compatible with either cognitivism or non-cognitivism being true. Conclusion (1), however, does not mean that normative propositions necessarily *can* be true or valid in any sense: one can argue that, while moral cognitivism is correct, *all* normative propositions are false. Furthermore, these three possible outcomes that I have mentioned, must, and will, be given further nuance. I am under no illusion that I will be able to *conclusively* show that any one of these positions ((1), (2), and (3), that is) is the case. Any conclusion that I reach is based upon the arguments provided in this thesis and due to the limited time and space of a master's thesis there are likely to be important aspects that has been overlooked.

1.5. Disposition

What follows this introductory chapter is an analysis of one of the early versions of Habermas's discourse theory of morality. The first section of this chapter begins with an analysis of Peter Strawson's essay 'Freedom and Resentment', which I decided to include in this analysis since Habermas provides an analysis of this essay himself in the beginning of his own essay, and it seems to me that Strawson's discussion can help to explain Habermas's cognitivism. I then turn my focus towards the two principles of which Habermas discourse theory of morality consists. I analyse and discuss the relation between them and some alternative interpretations. I also introduce the concept of communicative action and his view on argumentation and discourse and what this has to do with moral philosophy. The second section has a direct focus on understanding Habermas's cognitivism. As its title, 'Reason, Rightness, and Truth' suggests, the focus is on understanding the distinction between rightness and truth, and his concept of reason. Analysing Habermas's metaethics, I voice two concerns with his position that gives reason to doubt his claim that moral cognitivism is necessary for the rational justifiability of moral principles. The first of the two concerns Habermas's criticism of non-cognitivism,

especially the theory proposed by Hare, and the second focuses on criticism against Habermas's distinction between communicative and strategic action.

In chapter three, I attend to each concern over one section each, both having the purpose to provide an analysis that extends and elaborate the, up until this point, rather brief account of each issue. Firstly, I argue that Habermas's critique of Hare's prescriptivism is, while not without merit, rather unfair. The main point being that he misses important nuances in Hare's theory that contradicts the argument that Habermas tries to make, which might, in the end, defeat his claims. These nuances consist, among other things, of similarities between the two theories that Habermas does not account for in his criticism. I also consider some modifications of Hare's theory that is proposed by Grenholm, and whether they might help to decide the case. The second section of this chapter focuses on the distinction between communicative and strategic action which is central to his works in moral philosophy as well as epistemology and sociology. In this section I analyse three separate criticisms against this distinction and the effects of it for the discourse theory of morality. The critique is voiced by Benhabib, White and Dallmayr. Following these philosophers, I argue that Habermas is unable to sustain this distinction between communicative and strategic action and thus, that he is unable to justify the discourse theory of morality in the way that he tries to do.

In chapter four, I turn to the definition of cognitivism again and I try to give a more precise definition than I have done up to that point. While one may argue that it is problematic to have this chapter at the end of the thesis, due to it being concerned with the theoretical framework of the thesis, I think that my choice to put it where it is, is justified. The main reason being that much of the problems it addresses is better understood if those problems has already been explained. Some aspects relevant to this explanation will be given in the chapter focusing on the re authors, and also, it seems to me, is better dealt with in those chapters. In the first section, I will present my definition of cognitivism, explain when a statement is truth-apt and when it expresses a mental state like that of belief. In the second section I decide whether Habermas is or is not a cognitivist and why.

In the final chapter's first section, I once again turn to Habermas's critique of Hare. The purpose being to conclude the previous chapters' discussions of rationality, cognitivism, and truth-aptness. In the second section then, I conclude the discussions about the problems with Habermas's notion of communicative action and what consequences this has for his argument that cognitivism is necessary for rational justification. This chapter, in contrast to the others, does not end with a summary, but with a conclusion, a section where I do not merely summarise the discussions of the chapter, but also explicate what conclusions one might make.

2. The Discourse Theory of Morality

To conclude whether Habermas is correct in that cognitivism is necessary, I must explain several things; the most obvious being his cognitivist position and its function in his theory. To do this, however, I will not only analyse his cognitivism, but also give at least a general account of the discourse theory of morality to provide some context. I begin (2.1.) with the latter project focusing on the two principles of which his theory consists, and his attempt of a theory of justification. Thereafter (2.2.) I try to explain his metaethics.

2.1. The General Idea

In the essay 'Discourse Ethics: Notes on a Program of Philosophical Justification', Habermas takes his departure from Strawson's essay, 'Freedom and Resentment'. I will try to give a brief account of this analysis, since it has a central position for the thoughts that Habermas goes on to develop. The aspect of Strawson's essay that Habermas takes most interest in has to do with moral feelings and the observation that persons respond emotionally when harmed, and that this response takes a different form if the wrong was intentional than if accidental. Strawson writes:

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first.[14]

This difference is important, not only because it affects one's reaction, but also because that in realizing that unintentional harm has been committed, the perpetrator or a third person may apologise. The apology is important to the analysis since it shows, quite clearly, how one can change the other's view of the situation by giving reasons. By focusing on the *act*, one may try to add context to it so that it appears less wrong or unjust, or, one may focus on the *actor*, by urging the victim to view the perpetrator as incompetent.[15] In the latter case, the injury might have been caused by a child, or someone affected by alcohol, and while one might respond to them with anger, it will not evolve into resentment since they are not seen as being immoral or mean but merely immature or in some other sense incapable of acting rationally.

Habermas, following Strawson, calls this the objective, or third-person, perspective since it requires one to distance oneself from the situation. Taking this perspective, one cannot perceive the moral dimensions of interaction. Thus, while one can take this perspective as a '…

---

[14] Peter F. Strawson, 'Freedom and Resentment', in *Freedom and Resentment and Other Essays* (London: Methuen & Co. Ltd, 1974), pp. 1–25., p. 5.
[15] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 46.

refuge … from the strains of involvement … or simply out of intellectual curiosity',[16] it is not possible to completely step outside one's social world and remain there. To avoid the isolation that follows a consistent objective position towards interpersonal relationships, one must always return to the perspective of a participant in interaction. Thinking in similar ways, Habermas considers this an insight very relevant for moral philosophy. If one cannot grasp the moral dimension of an action from the objective perspective, then, neither can one locate the *moral point of view* outside the web of feelings and attitudes that makes up the social world. This is possible only from inside the performative attitude of someone participating in interaction.[17]

Returning to Strawson's example above, the resentment felt when having one's hand stepped on will be directed towards a specific person. This may be on behalf of oneself, or in the case of seeing someone having their hand stepped on, of that other person. The response itself, however, is not a reaction to the disturbance in the interaction, but to the violation of a general behavioural expectation. This might seem clearer from the perspective of the perpetrator: knowing that one has done something wrong one will feel guilty, a feeling that arises from disregarding an expectation that some action should or should not have been performed (not however, depending on its actual validity or invalidity). The same reasoning can be applied to feeling obliged to act in some way as well. Knowing about a behavioural expectation, and that it applies to the situation that one is currently in, one will have a sense of obligation to meet with it. Strawson suggests that these different categories of reactive attitudes are normally, to different degrees, evenly developed. One usually has similar normative expectations on others' behaviour towards others and oneself, as one's expectations on oneself towards others. Expectations that oneself as well as others are to be treated with a certain amount of good will, a claim to general validity which seems to be what gives a behavioural expectation its 'moral authority'.[18]

The important conclusion that Habermas draws from Strawson's analysis is that his observations shows that the behavioural expectations, rooted in the network of reactive feelings and attitudes in which we live our everyday lives must rely on a cognitive basis, and doing so they can be rationally justified. Feelings, he concludes this first section of his essay by saying,

---

[16] Peter F. Strawson, 'Freedom and Resentment', p. 9; Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 46, 47.

[17] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 47f.

[18] Peter F. Strawson, *Freedom and Resentment*, p. 15f; Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 48f.

seems to have the same function regarding the justification of claims to normative rightness, as sense perceptions has for the justification of truth-claims.[19]

I have tried my best to be brief in my above account of Habermas's interpretation of Strawson. I do believe, however, that what has been said above will help me to explain Habermas's metaethics and moral philosophy. The *discourse principle* affirms this belief, since it rests on the idea that part of a norm's authority, that is, its action coordinating force, is due to its general acceptance, in the sense that one does not only expect of someone else, or of all others to comply with it, but also feel obliged to do so oneself. To this insight, it adds a normative dimension stating that for a norm's validity to be at least considered in a practical discourse, it must not only be general, but also approved by those affected. It is formulated as follows:

> Only 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*.[20]

It is not, however, entirely clear what Habermas has in mind regarding the function of this principle. Towards the end of the essay, he states that it serves to define a formal procedure for testing whether norms can claim validity or not.[21] However, there seems to be a significant amount of overlap, both with the *universality principle*, which will be introduced in a moment, and the notion of *communicative action* which is central to both his cognitivism and moral philosophy. Acting communicatively, actors "… coordinate their action consensually, with the agreement reached at any point being evaluated in terms of the intersubjective recognition of validity claims".[22] Attempting to describe the usual mode of social interaction, this concept also carries with it a similar idea as the discourse principle does. As participants in practical discourse, their task is to rationally motivate the other participants using fallible reasons, to accept or reject a norm, a process of reason-giving that characterises the concept of communicative action.

The universality principle, in its turn, is inspired by Immanuel Kant. As formulated by him, the principle holds that one should only act according to such maxims that one would want to become universal law.[23] As Kant shows, with the example of someone asking to borrow money from some other person, the maxim cannot be universalised unless their intentions are

---

[19] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 49f.
[20] Jürgen Habermas, op.cit., p. 66.
[21] Jürgen Habermas, op.cit., p. 103.
[22] Jürgen Habermas, op.cit., p. 58.
[23] Immanuel Kant and others, *Groundwork for the Metaphysics of Morals* (New Hampton: Yale University Press, 2002)., p. 38.

honest. For Kant, lying violates the universality principle since one could not want a maxim that encourages lying to become universal law, since if that would be the case, the necessary preconditions for telling lies would not exist. The lie can only be successful if those involved assumes that *borrowing* money means giving the money back sometime in the future, and that everyone involved is sincere in their intention to do this. If these conditions are not the case, then one cannot expect the others to be truthful or keep their promises, and hence, no one will agree to borrow anyone any money since they would not expect them to be returned. In short, if universalised, such a maxim would be self-contradictory.[24] Turning my attention to Habermas's principle, a norm is valid if:

> all affected can accept the consequences and side effects that its 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).[25]

In this case, rather than being a formula against which one can test the validity of one's maxims, this version of the universality principle takes the form of a rule of argumentation. Rather than testing whether the norm, if universalised, contradicts itself, one should test its universality directly against the will of others.[26] In this way, Habermas suggests, the *general will* that the universality principle in its different incarnations[27] tries to represent, is more successfully represented by his version. In contrast to Kant's version of it, whose monological treatment of different perspectives includes the risk of unintentionally distorting them, the dialogical treatment of these interests that Habermas propose, does not pose the same risk, he claims. Instead, by making the individual the '… last court of appeal for judging what is in their own interest',[28] and allowing them to participate in the deliberations themselves, one avoids this problem.

Being a rule of argumentation, it is a principle with which any speaker participating in a practical discourse must necessarily comply. Anyone that does not follow this rule causes what is called a *performative contradiction*. As defined by Habermas, a performative contradiction means that '… a constative speech-act $k(p)$ rests on noncontingent presuppositions whose propositional content contradicts the asserted proposition p'.[29] To exemplify, again using the

---

[24] Immanuel Kant and others, *Groundwork for the Metaphysics of Morals*, p. 39.

[25] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 65.

[26] Jürgen Habermas, op.cit., p. 67.

[27] Jürgen Habermas, op.cit. p. 63. Here, Habermas suggests that every cognitivist morality relies on a different version of the universality principle. Whether this quite strong claim is true is beyond the scope of this thesis to discuss. The fact remains, however, that this principle has been formulated quite a few times throughout the history of modern moral and political philosophy.

[28] Jürgen Habermas, op.cit., p. 67.

[29] Jürgen Habermas, op.cit., p. 80.

liar used in Kant. In his example, *p* would refer to the proposition: 'By lying, I made them *agree* to lend me some money'; and the noncontingent presuppositions that agreements are entered into freely. There is a performative contradiction since coercion or manipulation of any form cannot constitute an agreement. As Habermas expresses himself:

> … everyone who seriously tries to *discursively* redeem normative claims to validity intuitively accepts procedural conditions that amount to implicitly acknowledging [the universality principle].[30]

These procedural conditions, or rules, are divided into three separate levels: the *logical-semantic* level, the *procedural* level, and the *process* level. The first level of rules determines the production of 'intrinsically cogent arguments' able to 'redeem and repudiate validity-claims'.[31] The second level consists of rules that defines the certain form of interaction that is argumentation; a procedure within which participants compete for the better argument.[32] The third level regulates the communicative process itself, determining the conditions for participation. This is what Habermas had earlier described as the 'ideal speech situation'. As such, its premises are supposed to create a process free of any form of force other than that of the better argument. It is the second and the third level that is of interest here since they focus on the relations between the participants, eliminating any relations of domination between them.[33] On the procedural level, Habermas gives two examples of rules, the first specifies truthfulness and says that 'every speaker may assert only what he really believes', another rule from this level says that 'every person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so'.[34] The processual level of rules specifies, among other things, that '[e]very subject with the competence to speak and act is allowed to take part in a discourse',[35] and that anyone allowed to participate may introduce or question any assertion whatever, and express their attitudes, desires, and needs.

The rules located in this final level, the process-level, are the most significant ones. As essentially consensus-oriented, actors cannot violate the rules of this level, nor those of the second one, without getting themselves ensnared in a performative contradiction.[36] Since reaching an agreement means that the participants have freely accepted some norm or set of

---

[30] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 92f.
[31] Jürgen Habermas, op.cit., p. 87.
[32] Jürgen Habermas, op.cit., p. 87.
[33] Jürgen Elfabermas, op.cit., p. 88f.
[34] Jürgen Habermas, op.cit., p. 88.
[35] Jürgen Habermas, op.cit., p. 89.
[36] Jürgen Habermas, op.cit., p. 87.

norms as valid, they have committed themselves to act accordingly.[37] As Habermas explains, consensus is formed by participants *convincing* their opponents that their claims are the more sensible ones. If a rule from either the second or the third level are broken, say, that they exclude someone from participating in the discussion or that they lie to them (as in my previous example), and only then reach an agreement or convince someone of something. In that case the conviction or agreement has been formed '... under conditions that simply do not permit the formation of convictions' [38] and agreements and thus cause a performative contradiction.

What confuses me a bit is that all these rules of the procedural and the process level (which all implies the universality principle), also seems to promote the same idea that underlies the discourse principle, that is, to frame social interaction as a communicative process. That everyone affected must be included in the discussion, as the discourse principle states, is stated both in the formulation of the universality principle itself, as well as in the third level of rules derived from it. Applying this principle on argumentation, as Habermas does, it seems like the discourse principle becomes superfluous. Habermas seems to be aware of this as a source of confusion. He states explicitly that they are *not* two different formulations of the same principle, and that '… [w]e should not mistake this principle of universalization for [the discourse principle], which already contains the distinctive idea of an ethics of discourse'.[39]

Starting with the differences, the discourse principle only states that a norm *can claim* to be valid, if it meets, or hypothetically could meet, with the approval of those affected in their capacity *as participants in a practical discourse*. This principle opens for several possibly conflicting norms competing for the participants' approval. Excluding from the discussion, any norms that are, and would be, unable to meet with their approval. The *universality principle* then, states that a norm *is* valid if those affected can accept *the consequences of its general compliance*, excluding instead, all norms that has not been explicitly accepted by the participants.

There is also another interpretation of the relation between the two principles, quite possibly it is also a more plausible one. There are indications in the text, some quite strong one's, that Habermas intends the universality principle to be the fundament on which the discourse principle stands, which is probably the correct interpretation. Framing the universality principle as a "bridging principle" which is to fill the logical gap between observations of moral

---

[37] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 59.
[38] Jürgen Habermas, op.cit., p. 90.
[39] Jürgen Habermas, op.cit., p. 66.

15

Johan Elfström

phenomena and generalised norms is one such indication.[40] Claiming that the discourse principle already assumes that normative expectations *can* be justified, that the justification of the universality principle is the most promising 'road' towards the justification of the discourse principle,[41] strengthens this interpretation.

The problem with this interpretation is that it would render the discourse principle even more superfluous, and I cannot simply see that it would have any function at all to fill in his theory. This, I argue, is because it is so much less specific than the universality principle. As the latter states that a norm *is* valid if everyone affected accepts its consequences and prefers them to the alternatives, there is no need for a principle that suggests that a norm *can claim* to be valid if it can be approved of in practical discourse. It does indeed specify the context to that of practical discourse, but that, or at least the context of argumentation, is already presupposed in the rules of argumentation form which Habermas claim to derive the universality principle. Seyla Benhabib has argued differently and claims instead that it is the universality principle that becomes redundant in the light of the discourse principle. Her reason is that the universality principle is inferior to the discourse principle and adds nothing but 'consequentialist confusion'.[42] That everyone affected must be able to accept the consequences of the general compliance, she claims, abandons Kant's ideal of respecting the autonomy of others. Thus, her problem with Habermas's formulation is that it strives to guarantee consensus on moral issues, which, according to Benhabib, does not have moral value in itself.

2.2. Reason, Rightness, and Truth

Proceeding from the vantage point that the previous section provides, the focus of this section is those aspects of Habermas's metaethics that are necessary in order to understand what he means with cognitivism, and his claim that cognitivism is necessary for the rational justification of moral norms. He suggests that sceptic critique of moral cognitivism: '… relies primarily on two arguments: first, the fact that disputes about basic moral principles ordinarily do not issue in agreement, and second, the failure … of all attempts to explain what it might mean for normative propositions to be true'.[43] Against the first one, Habermas tries to show that agreement can be made possible, and proposes the universality principle as a solution. Against the second claim, he suggests that the '… premise that normative statements can be valid or

---

[40] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 63.
[41] Jürgen Habermas, op.cit., p. 94.
[42] Seyla Benhabib, *Situating the Self*, p. 37.
[43] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 56.

invalid only in the sense of propositional truth'[44] should be abandoned, and instead, that they should be considered analogous.

The notion of validity claims is central to these tasks. Developed from previous phenomenological theories, which did not reach beyond the objective world, Habermas notices that '... the rationality of persons are not exhibited solely by the ability to utter well-grounded factual beliefs and act efficiently',[45] and thus attempts to extend this analysis to other forms of claims too. As Habermas recognises, raising validity-claims, speakers relate their statements not only to an objective world, but a social, and subjective world too. Claims to *truth* are related to states of affairs in the *objective* world; *rightness-claims* relates to the web of behavioural expectations that is the *social* world, and *sincerity-claims* relates to the *subjective* world. In all cases, the speaker can, if asked, justify their claims to validity by providing fallible reasons in their defence. In the case of sincerity, however, since the speaker have privileged access to the world to which such claims are related, such claims cannot be justified in this way, but must be assessed in the light of the speaker's behaviour. What this comes down to, is that through engaging in discourses, the objective and the social world alike, acquires objectivity in the sense of starting to count as one and the same for the persons belonging to a certain communication community.[46]

Now, it should be noted that Habermas does not seek to deny nor question the existence of a mind independent objective world. He does actually make an important point about the objective world having an existence independent of the claims to validity related to it, which the social world has not. Were they mind independent objects, they would be unable to coordinate behaviour in the sense they do.[47] A normative statement, say, 'one ought not to lie', differs from a purely factual one, such as 'the computer is broken', in that the latter is meaningful only when raised. In case of this first statement, verbally articulating the norm 'one ought not to lie', does not add to it any meaning that it did not have before it was uttered, but in the latter case the meaning of the proposition rests entirely in the speech-act. Assume that the assertion is made as a response to someone asking a question; whether they may use the computer. That it is a response to this question is an important part of the meaning of the utterance. Apart from merely stating that the computer will not work properly or not even start; as a response to the question it has an action coordinating effect. In order to carry out their goal,

---

[44] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 56.
[45] Jürgen Habermas, *The Theory of Communicative Action. Vol. 1, Reason and the Rationalization of Society* (Boston: Beacon Press, 1984)., p. 15.
[46] Jürgen Habermas, *The Theory of Communicative Action*, p. 12ff.
[47] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 60.

they must use another computer, repair it, or something in that direction. Norms, on the other hand, exist because agents believe, not only that others will comply with them, but also that it is *expected* of themselves as well as of others to do so. Its action coordinating effect is intrinsic to the norm and is not added by the speech-act in which it is uttered as is the case with truth-claims.

Indeed, the existence of a norm, or the expectancy to comply with it, does not mean that it *is* valid, and thus, that it is worthy of recognition. But, the feelings that may give rise to normative claims to validity seems to make up the basis of moral-practical discourse, in which these claims can then be redeemed or rejected. Here, as has been shown, Habermas proposes the universality principle, claiming it to have a similar function to the principle of induction in theoretical discourse, which is to '... bridge the logical gap between particular observations and general hypotheses'.[48] As the principle of induction serves to make agreement possible about factual matters by turning *our* repeated interaction with the objective world, into general hypotheses, so the universality principle does with the participants' experiences of interaction in the social world.

This breaks abruptly with traditional analyses of social interaction, which has been restricted to the objective world: action has been analysed in terms of purposive-rational action performed by actors oriented towards their own success; communication and meaning has similarly been located in the speaker's intention. Habermas thinks that these analyses are flawed since they obscure important aspects of interaction for actors and social scientists alike (see the analysis of Strawson in the previous section). Consequently, they are not merely inadequate theories for understanding interaction, but they give interaction an unfortunate strategic touch.[49] If someone were to tell their friend that 'I feel sad today', it seems odd to claim that the hearer has understood the meaning of what has been said if they have understood the intention behind the speaker telling them about their feelings. It does not seem to be necessary that the speaker intends anything else than to communicate to their friend how they are feeling. If that would be the case, it seems rather absurd to say that one understands the meaning of a statement if one knows the intention behind it. The hearer will not be able to infer what the speaker means by saying "I am feeling sad" from the knowledge that they seek to establish a definition of the situation from which the two can continue to interact with each other.[50]

---

[48] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 63.
[49] Jürgen Habermas, *The Theory of Communicative Action*, p. 95.
[50] Jürgen Habermas, op.cit., p. 275.

Rejecting the intentionalist theory, Habermas proposes what is called truth-conditional semantics, according to which the hearer understands a statement when they know under what conditions it is true. This theory does indeed come with a similar problem as the intentionalist one: it is counterintuitive to claim that 'it is *true* that I feel sad'. The problem being that this theory, as the intentionalist one, does not make any effort to reach beyond the objective world. This problem can however be resolved through a modification: by extending the conditions according to which the hearer is thought to understand a statement from the rather narrow concept of truth to the more general concept of validity.[51] This modification, the move from truth to validity conditions places the knowledge of what a statement means in understanding under what circumstances one could accept the conditions specified. In the example above, the speaker reveals their emotional state and places the validity conditions in what might cause such a state and the speaker's past and future behaviour. Raising a claim to normative rightness then, the validity conditions resides here too, on the one hand, in the propositional content, and on the other, in the obligations which arise for the hearer were they to accept the claim.[52]

This theory of meaning is, in the theory of communicative action, combined with speech-act theory. As Habermas explains, a speech-act always consists of a statement in the form of a sentence and its 'mode'. When someone says to their friend that they are sad, there is first an utterance, the verbal articulation of the sentence; then there is the 'doing' in it, making a promise, asking a question and the like. Both are necessary aspects in any utterance whether one states some fact or raises a claim to rightness or sincerity. Having extended the idea of truth conditions to validity conditions, the meaning of the utterance is located not only in the sentence itself, its content so to speak, but also in the 'doing'. The hearer knows the meaning of what is being said if they know the conditions for being sad, and who the 'I' refers to. In other words, I will assume, they have understood what the speaker means if they know that the speaker is referring to an emotional state of theirs, what it is like to have such feelings. Regarding the doing, in Habermas's example, he calls this form of speech act a confession, which seems to be appropriate in my example too; the speaker confesses to the hearer what their emotional state is. The hearer then, understands the meaning of this confession if they know how this feeling will manifest itself in the speaker's behaviour.[53]

This account of Habermas's philosophy of language provides much to the understanding of the rationality of moral argumentation – the use of criticisable reasons to motivate hearers to

---

[51] Jürgen Habermas, *The Theory of Communicative Action*, p. 277f.
[52] Jürgen Habermas, op.cit., p. 302f.
[53] Jürgen Habermas, op.cit., p. 296f, 303.

accept or reject claims to normative validity – which is central to his critique of non-cognitivist theories such as emotivism and prescriptivism.[54] Habermas explains that these theories claim that normative propositions merely express one's subjective preference for or against some action and invites others to share them. This means that they are not connected to fallible claims to validity that can be criticised, accepted, or rejected by participants in practical discourse.[55] While one may, of course, challenge the rationality of someone's feelings towards an action by criticising the sincerity of it, that misses the point. A normative statement like 'one ought not to lie', does not, at least not for emotivists like Alfred Ayer,[56] mean that the speaker *asserts* their dislike of lying. Rather, Grenholm explains, it should be understood in the sense of the statement mediating, from the speaker to the hearer, the former's negative emotions towards lying and their wish for the other to feel the same way.

Regarding prescriptivism, Habermas acknowledges that Hare's theory does take moral argumentation into account, which does not, however, change his conviction. He claims that it is a form of ethical decisionism, since one may arbitrarily choose principles to guide one's decisions in practical matters.[57] Seemingly agreeing with Habermas on this matter, White describes Habermas's cognitivism as:

> […] one which follows the Kantian tradition in arguing, first, that valid norms are one's which have the quality of fairness or impartiality; secondly, that this quality can be expressed by some version of the principle of universalization; and finally, that this principle itself can be rationally justified.[58]

The non-cognitivist, he then claims, would disagree with the last statement on the grounds that the choice to comply with such a principle is but a decision, meaning that one is not rationally bound to accept the universality principle. In contrast, Grenholm provides a more generous interpretation of Hare's prescriptivism. While Habermas concludes that the choice of moral principles, according to this theory, cannot be justified, Grenholm does not seem to agree with this. Contrary to Habermas's conclusion, he claims that Hare does speak of rational justification of moral judgements and principles,[59] and that a variation of the universality

---

[54] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 54f. Habermas seems to use the word imperativism as a general category to which prescriptivism is belongs. Since I will not discuss the general theory, that is, imperativism but only prescriptivism, I will use that word consistently to avoid confusion.

[55] Carl-Henric Grenholm, *Etisk teori*, p. 56; Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 54, 55.

[56] Carl-Henric Grenholm, *Etisk Teori*, p. 53.

[57] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 55.

[58] Stephen White, *The Recent Work of Jürgen Habermas: Reason, Justice and Modernity* (Cambridge: Cambridge University Press, 1988)., p. 48.

[59] Carl-Henric Grenholm, *Etisk Teori*, p. 78ff.

principle is central to this. If Grenholm is correct, then this cannot be what distinguishes cognitivists from non-cognitivists. He claims that even to Hare, the universality principle makes rational argumentation on moral issues possible, and more importantly, universalizability is a criterion for an acceptable moral judgement. Thus, while the rational justification of the universality principle is important to Habermas's cognitivism, it is not what characterises it, nor cognitivism in general.

Indeed, as has been shown, Habermas is aware of this aspect of Hare's philosophy, and his argument against Hare's use of the principle suggests that it works on a different level in Hare's theory than it does in Habermas's. As Habermas tells it, Hare claims that the universality principle is a logical consequence of semantic rules, and that it therefore is just as inconsistent to assign different normative predicates – that something is right or wrong – to instances that are similar in all relevant regards, as it would be to do this with descriptive predicates.[60] In other words, Hare restricts his version of the principle to the logical-semantic level of argumentation, and is thus too thin to be a moral principle since one could use it to justify a norm that is deeply immoral on the grounds that oneself would accept its universalisation. In order to be a moral principle it must include a notion of impartiality, among other things. But, as Habermas concludes: '... one cannot derive the meaning of impartiality from consistent language use'.[61] If Grenholm is correct in that prescriptivism allows for rational justification of moral norms and principles however, then this aspect of Habermas's argument has no impact on the rational justifiability of moral norms and principles.

The main problem to Habermas is that non-cognitivist theories takes an 'objectivating attitude'.[62] As already established, Habermas agrees with Strawson that one cannot grasp the moral dimension of social interaction from outside the participant's perspective. As with emotivism, Habermas claims that Hare reduce the validity of moral judgements to merely universalisable personal wishes as moral statements are expressed in the form of imperatives. Imperatives, as opposed to validity-claims, are, according to Habermas, demonstrations of force and does not have the reflective rational potential that Habermas sees in validity-claims. The hearer has understood an imperative if they have understood what the speaker says, and thus what the speaker desires, and what would bring the desired state into being. In this sense they do not differ much from other expressions. However, an imperative request does not appeal to

---

[60] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 64.
[61] Jürgen Habermas, op.cit., p. 64.
[62] Jürgen Habermas, op.cit., p. 55f.

the validity of norms but will only be given in to if the speaker by means of sanctions or rewards can influence the hearer to comply.[63]

Considering validity claims then, in the reason-giving process of communicative action, participants are rational if they raise claims to validity which are intersubjectively recognisable. This intersubjectivity of reason is important for the action coordinating function of discourse, theoretical and practical alike. In raising a validity-claim, there is also an underlying promise that the speaker can, and will, provide reasons for the validity of their claim, were it to be challenged.[64] If the participants engaged in practical discourse agree to accept a norm about not lying to others as valid, then there also follows a symmetrical obligation on everyone involved in the discourse to comply with the norm whose validity they have agreed to. If there was no guarantee that a claim to validity could be redeemed by providing reasons in its favour, neither would there be an obligation to comply with the norm. This is because if one could not provide any reasons, or if one were to refuse to do so, then the only way to enforce the compliance of the rule is through force, and the legitimacy that follows rational agreement no longer exists. The same way it is with the need for the validity-claims to be intersubjectively recognisable: if one were to, say, express oneself in an idiosyncratic fashion, so that no one else could assess the validity of one's reasons, then communication, as understood in the context of Habermas's action theory fails in a similar way.

This is the core of Habermas's claim that there is need for cognitivism. He summarises his critique of the different non-cognitivist positions in the following way:

> None of these types of sentences [experiential, imperative, intentional] can serve as a vehicle for making a truth claim or for making any claim to validity that requires argumentation ... In short, with a single blow noncognitivist approaches deprives the sphere of everyday moral intuitions of its significance.

There are important similarities between the distinction Strawson makes between the participant's perspective and the deviant 'third-person observer', and Habermas distinction between communicative and strategic action. The strategic actor, for instance, while not necessarily ignorant of anyone's feelings and interests, will act in disregard of them with the purpose of achieving their own individual goal. This argument creates some problems for Habermas. It seems open to the possibility to choose not to act communicatively, and thus escape the burden of the rules of discourse entirely. Habermas would reply that since our reactive feelings are rooted in the framework that is our social world, if one were to retract

---

[63] Jürgen Habermas, *The Theory of Communicative Action*, p. 298f, 301.
[64] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 58f.

oneself from this framework for any longer time, one would be unable to sustain any interpersonal relationships or engage in any interaction at all. Social interaction is, according to Habermas, necessarily communicative.

I do believe that Habermas is on to something here. That communicative action, or something like it that might warrant similar presuppositions of honest and rational actors, lays the foundations of a stable society does seem to have merit. If the norm was dishonesty, as an example, it would be unlikely for stable relations to form. Nonetheless, if one were to follow Habermas in this, and claim that communicative action is necessary as an argument against scepticism, the sceptic can simply choose not to engage in the process of argumentation and thereby prove their point. If communicative action is necessary, it is odd that one can stop acting communicatively. To such an argument, Habermas replies that:

> By refusing to argue, for instance, he cannot, not even indirectly, deny that he moves in a shared sociocultural form of life, that he grew up in a web of communicative action, and that he reproduces his life in that web … In other words, he cannot extricate himself from the communicative practice of everyday life in which he is continually forced to take a position by responding yes or no. As long as he is still alive *at all*, a Robinson Crusoe existence through which the skeptic demonstrates mutely and impressively that he has dropped out of communicative action is inconceivable, even as a thought experiment.[65]

There is only an abstract possibility to choose between communicative and strategic action; it exists '... only for someone who takes the contingent perspective of an individual actor',[66] but not from the perspective of the social world in which they live their everyday lives. The argument is convincing, I would say, in the sense that it successfully shows that one cannot have a meaningful social existence if one were to consistently refuse to act communicatively. What Habermas fails to show, however, is that one is rationally obliged to *always* act communicatively, which is, I assume, the conclusion that he wants his argument to entail. The obvious objection to my doubts is that someone who sometimes choose to act communicatively and sometimes not, behaves inconsistently and thus irrationally. This is also what Habermas is saying; that the sceptic is forced, either to accept that he must act communicatively, or completely drop out of the web of communicative action, which simply is not possible. However, even if one cannot maintain a consistently objectivating attitude towards other actors, it does not appear to be rationally forbidden to take this attitude only sometimes. One might even be justified in doing so in some instances. Habermas's explanation of Strawson's analysis

---

[65] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 100.
[66] Jürgen Habermas, op.cit., p. 102.

of apologies provides one example where forgiveness can be granted, *only* because one takes an objectivating attitude towards the actor responsible for the wrong. In Strawson's example, since the perpetrator was not a competent actor, for some reason – young age, stressed out, or affected by alcohol – by taking an objectivating attitude towards them, one can dismiss the seriousness of their action. To forbid rational actors from sometimes taking an objectivating perspective, thus seems counterintuitive.

When I let White describe Habermas's cognitivism above, he begins an argument which eventually ends up in a place similar to this one. As he argues, Habermas is correct in his claim that the "speech-act immanent" obligation to stand ready to justify a validity-claim cannot be entirely avoided.[67] It does not seem like anyone can live an entire life outside the web of communicative action and thereby avoid finding themselves in situations where they must, seriously and sincerely, defend or challenge a claim to validity. The problem for White however, is that an actor is not rationally bound to accept the universality principle that Habermas propose, but could instead adopt a less reflective moral principle. As long as it does not violate the obligation to stand ready to justify one's validity-claims, the actor will not get entangled in a performative contradiction by adopting it. The conclusion that White draws is that Habermas might not have been able to show that the version of the universality principle which he proposes must be accepted by any rational agent, although he has been able to sufficiently argue in favour of rejecting value-scepticism.[68]

2.3. Summary

In this chapter I have analysed Habermas's discourse theory of morality, and the cognitivist meta-ethical position he elaborates as it is explained in the book *Moral Communication and Communicative Action* and in *The Theory of Communicative Action*. In the first section, I provide an interpretation of Habermas's moral philosophy in order to give the relevant context to the metaethical aspects of it. The main problematic being the relation between the two principles of which his theory consists, settling in the end of that section with an interpretation of the universality principle as the more basic principle on which the discourse principle rests. The reason for this is that, as Habermas says, the discourse principle already assumes that norms can be justified, therefore, to be able to apply this principle it must be shown that one might actually do so.

---

[67] Stephen White, *The Recent Work of Jürgen Habermas*, p. 55.
[68] Stephen White, *The Recent Work of Jürgen Habermas*, p. 58.

In the second section, besides analysing Habermas's metaethical position, I point out two problems with Habermas's argument in this analysis. The first one concerns the claim that the universality principle is a characteristic of Habermas's cognitivism, I point out that Hare's and Grenholm's prescriptivist and thereby non-cognitivist approach to metaethics also includes a version of the universality principle. While there are important differences between the prescriptivist version of the principle, and between Habermas's account of it – the prescriptivist one is for instance not a *moral* principle but is merely a consequence of semantics and logic – these differences do not seem to be of such importance so that they render rational argumentation on moral issues impossible. As for the second problem, it concerns the distinction between strategic and communicative action and its corresponding distinction between the participant's perspective and that of the third-person observer. To be able to sustain his strong claim about the necessity of cognitivism, Habermas must be able to show that anyone engaging in social interaction cannot escape the rules of discourse and therefore, that they must accept the universality principle. With the help of White, I have given some reasons to doubt that this is the case.

3. The Two Problems Elaborated

In this chapter, I will elaborate on the two problems of Habermas's cognitivist position which I have tried to point out in the previous chapter. I begin my analysis (3.1.), in the critique directed by Habermas against prescriptivism, and delve deeper into the question of the merits of Habermas's criticism of non-cognitivism. Then (3.2.), I consider whether the distinction between communicative and strategic action, and the corresponding distinction between the participant's and the observer's perspective is as problematic as it seemed in the previous chapter.

3.1. On the Universality Principle

In the previous chapter, I made two arguments which indicates that Habermas's claim, that prescriptivism, as a form of non-cognitivism, cannot justify moral norms is problematic. In this section, I will focus on the plausibility of one of those arguments; that the absence of a moral principle does not, contrary to what Habermas suggests, seem fatal to the prospects of. I have already made a brief analysis of Habermas's views on this matter, and thereby, I have been able to indicate that some of his claims seems incorrect. Here, I will give much closer attention to Hare's and to Grenholm's versions of prescriptivism and consider them from the perspective of Habermas's criticism.

Contrary to Habermas, according to whom the principle is analogous to inductive reasoning, Hare finds a similarity with the hypothetical-deductive method. The idea is that a moral judgement is justified if its consequences remains acceptable when the judgement is universalised. Say that one condemns someone for telling a lie, that person is justified in doing so if they can accept the consequences it would have in any identical situation, regardless of whom the person telling the lie is and who gets lied to, and so forth.[69] Consider Kant's example with the liar that I previously mentioned. Assume that the law (the positive law, not the moral one) states that '... creditors may exact their debts by putting their debtors into prison'.[70] Assume too, that the person lied to in Kant's example (I follow Hare and call this person 'B'), now considers doing just that to the person who borrowed the money (let me call this one 'A'). Considering whether this is an adequate measure to take, B realises that they too owe money to someone else (person 'C', that is). If deciding that they, B, *ought* to put A into prison; following the universality principle, they are, thereby saying that C ought to put them in jail too, and that

---

[69] Carl-Henric Grenholm, *Etisk teori*, p. 80.
[70] Richard Hare, *Freedom and Reason* (London: Clarendon Press, 1965)., p. 90f.

they should let them do so. Assuming then, that B is not inclined to go to prison, and thus, considers the consequences of the universalisation of the ought to be unacceptable, B is not justified in putting 'A' into prison either. Hare continues:

> We must first notice an analogy between [the universality principle] and the Popperian theory of scientific method ... What knocks out a suggested hypothesis, on Popper's theory, is a singular statement of fact: the hypothesis has the consequence that p; but not-p. Here the logic is just the same, except that in place of the observation-statements 'p' and 'not-p' we have the singular prescriptions 'Let C put B into prison for debt' and its contradictory. Nevertheless, given that B is disposed to reject the first of these prescriptions, the argument against him is just as cogent as in the scientific case.[71]

In other words: Hare is interested in the testing of the hypothesis, whereas Habermas is interested in its formulation. This example also shows another important difference. For Habermas, the universality principle is not so much about whether B can justify putting A into prison by considering whether they would accept this prescription if it was issued against them. Rather than focusing on whether one may take advantage of a legal norm, Habermas's focus is on whether this legal norm, that one may exact debt by putting persons in prison, is justified in the first place. So, while an acceptable moral judgement must be universal according to Hare, it need not be general. That a speaker says to a hearer that "you ought not to have lied to them" does not imply the *general* statement that 'you ought not to lie'. The logics of *ought* does however mean that 'you' in the speaker's utterance cannot merely refer to the specific person to whom the utterance is directed.[72] Grenholm suggests that this means that it is possible, within the bounds of Hare's theory, to reject the existence of such general principles altogether (not that there *cannot* be any general principles, but that there *need not* be any). Hare regards reality as too complicated to formulate highly general moral principles as philosophers are often tempted to do. This being the case, attempting to formulate such principles, they will either be too simple to do their job very well, or, one will not be able to formulate any principles with the desired degree of generality.[73] In any case, the project is in vain.

It seems that one cannot reject generality entirely, however. Since if each situation is unique then one cannot learn from previous experiences.[74] There might be some situations that

---

[71] Richard Hare, *Freedom and Reason*, p. 91f.
[72] Carl-Henric Grenholm, *Etisk teori*, p.78; Richard Hare, *Freedom and Reason*, p. 31f, 38f. Now, it seems reasonable too, to expect that in a universal prescription of this sort 'them' does not refer to some specific person or persons but to a certain relationship, say that 'one ought not to lie to one's spouse (or the police or one's friends etc.)', phrased differently it might instead refer to a certain situation or a certain matter about which one ought not to lie. Indeed, it does not seem to be wrong in any logical sense to make a universal statement in the sense of 'one ought not to lie to this specific individual', although that would not be very helpful.
[73] Richard Hare, *Freedom and Reason*, p. 37f.
[74] Richard Hare, *The Language of Morals,* p. 61.

are this complex, but in most instances, one might be able to find some aspects that supports at least some degree of generality. For Habermas, as has been shown, generality has a much more prominent role to play, and it is the consequences of the *general* compliance of a norm that those affected by it must be able to accept it for it to be valid. While, in contrast, Hare equates generality with simplicity. The more general a principle is, the more abstract it must be, and consequently, it cannot take into consideration every complication one might encounter in real-life situations. Anyone with any decent amount of experience of social life will know that any principles one might hold *prima facie* might have to be revised facing a new situation. Since what characterise a genuine moral judgement is willingness to act on it, principles that are strictly general are thus unlikely to be genuine.[75]

To a certain extent, one might think that Habermas would agree. When Hare speaks about moral judgements, he is not very far away from speaking of what Habermas calls sincerity-claims. When someone makes a value judgement, about putting A in prison for instance, they express their subjective view on this matter. The comparison is further sustained by Hare considering it hypocritical to make judgements one does not intend to live by oneself. As he suggests, an actor can test the validity of a judgement considering whether they would accept the consequences when in the position of the affected. If they would not, they are not justified in acting towards someone else in this manner either. Up to this point, the comparison seems to work, and supports the claim that rational justification of norms and principles can be rationally justified even by a non-cognitivist theory. However, as is the case with sincerity-claims, since the speaker has privileged access to their subjective world, a hearer must evaluate the sincerity of the judgement against the speaker's behaviour and other judgements that they make.

This comparison reveals some problems for Hare. The subjective nature of one's judgements does not fit well with the idea of a speaker rationally motivating a hearer to accept a judgement as valid as they may hold contradictory judgements without one of them being correct and the other incorrect. Hare makes this position quite explicit, as he writes that:

> ... if pressed to justify a decision completely, we have to give a complete specification of the way of life which it is a part. This complete specification it is impossible in practise to give; the nearest attempts are those given by the great religions, especially those which can point to historical persons who carried out the way of life in practice. Suppose, however, that we can give it. If the inquirer still goes on asking 'But why *should* I live like that?' then there is no further answer to give him ... We can only ask him to make up

---

[75] Richard Hare, *Freedom and Reason*, p. 42f.

his mind which way he ought to live; for in the end everything rests upon such a decision of prinicple.[76]

It is against the background of this statement that Habermas concludes that according to Hare, rational justification of moral norms is impossible. Since one's judgements are a product of this principle, which is something each and every one must decide for themselves. At the same time, if one can make the above comparison, it becomes difficult to reject that Hare's theory allows at least some possibilities of justifying one's judgements, although, from Habermas point of view, wrongly views them as sincerity-claims rather than claims to rightness.

It counts against the overall comparison that Habermas, contrary to Hare, does not think that one is less rational if one's actions fail, or one's propositions are shown to be false, not as long as the belief motivating the action or proposition is otherwise carefully constructed.[77] Grenholm raises a similar objection: it does not fit well with our moral experience, he says, that one must succeed in following one's prescriptions to be sincere in expressing them. It is not unusual to do just that, that is, make a judgement that, to one's own despair, one fails to comply with oneself. Someone's rationality might be doubted, according to Habermas, it seems, if one seems unaware of oneself contradicting one's judgements, or similarly make judgements which has not been thought through, but also if one's lack of commitment to one's prescriptions are strategically motivated.

However, even if this comparison was to fail, as I have defined the notion of rational justification in the introduction, Hare's theory of moral reasoning does seem to meet the conditions set out there. As has been seen, Hare allows valid inferences between normative reasons and normative claims. Furthermore, having different conceptions of rationality, Habermas and Hare does actually agree on the basic distinction between rational argumentation and the use of coercion or manipulation to make others do one's bidding.[78] The way in which Hare's universality principle works, is that words with an evaluative function, words that is used to praise or condemn, such as good or bad and right or wrong, also have a descriptive function.[79] That there are reasons that motivates the commendation or condemnation of an action does indeed speak in favour of rational justification being possible. This was also mentioned as a condition to be able to rationally justify a claim.

---

[76] Richard Hare, *The Language of Morals*, p. 69.
[77] Jürgen Habermas, *The Theory of Communicative Action*, 11. Habermas says this with regards to goal-oriented action, although I am unable to see why that would not be the case regarding sincerity-claims too; it does not have to be the case that one's judgement is less sincere only because one has failed to act on it oneself.
[78] Richard Hare, *The Language of Morals*, p. 15.
[79] Carl-Henric Grenholm, *Etisk teori*, p. 75; Richard Hare, *The Language of Morals*, p. 80f, 111, 145, 153.

It is a reasonable conclusion to make, I believe, that in requiring universality, Hare's principle also reaches beyond the strategic mode since one might be required to abstain from pursuing a goal because it cannot be universalised. Consider the creditor seeking to exact a debt, they are rationally obliged to abstain from this whish or, at least, to exact the debt by some other means. This position is complicated by another aspect of Hare's theory that connects to the subjective nature he ascribes to moral judgements. Hare claims that there is not, in fact, that there *cannot be*, a logical connection in the sense of entailment between the reasons – the descriptive content, that is – and the judgement. If there was such a relationship, value-words would be unusable. Assume that there is the picture P which has a certain set of characteristics X and someone thinks that they should commend P for having X. If, however, 'P is a good picture' means *the same* as 'P is a picture and P has X' they cannot do that. By saying that 'P is a good picture because P has X', would in that case, be equivalent to saying 'P has X because P has X'.[80] The same problem arises to the other words in the moral vocabulary such as 'ought' and 'right'.[81]

To be explicit, what this means is that one cannot assume a set of premises and deduce from those premises the goodness or rightness of something unless at least one of the premises contains a value word in the first place. This may be used as an objection: one might say that if there is no relation of entailment between the descriptive and evaluative content of the value-word, rational argument about moral issues becomes impossible. If two persons disagree on a moral issue, and there is no such logical relation, then there is no way for the one to *convince* the other that they should change their mind. This objection fails to the degree that no *relation of entailment* does not imply that there is *no relation at all*, and the inductive relation (or at least something similar to an inductive relation) that is proposed by Habermas is one possible solution. Hare's answer, I believe would be similar. As Hare describes this aspect of his theory, one must simply assume, from one's general experience of human beings, what their desires and aversions are.[82] The logics does not enter until a later stage in the procedure. Recall the introductory example where person B thinks that they ought to put A in prison in order to exact a debt but does not think that C ought to put them, B, in prison for the same reason. Considering B's prescription, it follows logically from the universality principle that if C wishes to exact B's debt, B would be logically inconsistent unless they allow C to put them in prison too. If B is not inclined to go to prison, then they may not make their original prescription.

---

[80] Richard Hare, *The Language of Morals*, p. 85.
[81] Richard Hare, op.cit., p. 155.
[82] Richard Hare, *Moral Thinking*, p. 13f.

It is important to note that it does not matter whether or not B is actually indebted to C. It would certainly be a logically valid to simply change the premises along with the conclusion so that since B is not indebted to C, C ought not to put B in prison. However, as Hare would put it, in this case, one has misunderstood the universal character of 'ought'. The particulars of the situation are not relevant. If B claims that they ought to put A in prison to exact their debt, they say, at the same time, that were they ever in a similar position, that is being indebted to someone, then they assent to the same thing being done to them. It does not matter whether or not they are or will ever be in that position themselves, they must still test their prescription against their own inclinations and desires. This too speaks in favour of the rational justifiability of moral judgements considering that it illustrates my previous point Hare's theory allows valid inferences using normative reasons.

However, according to Habermas, there is a problem that the judgement is arbitrarily adopted by the speaker. This objection is only partially successful. Habermas is correct to the degree that the judgement, valid or not, comes close to being merely a subjective opinion. This has already been argued. Nonetheless, while Hare does indeed criticise moral relativists, he does not criticise them for granting someone else to hold different moral views than oneself. He criticises relativists for being self-contradictory. If one says that 'I ought not to lie' but does not care to universalise the 'ought' to apply everyone else too, then one has not understood the logics of the word 'ought'.[83] What supports Habermas criticism that moral principles can, in Hare's view, be arbitrarily adopted by the individual is that he denies that moral judgements can be decided to be correct or incorrect. Philosophers that make such assumptions, Hare claims, confuses moral philosophy with empirical sciences.[84] This aversion towards moral intuitions is a bit odd, considering that the meaning and use of value words, and thus, their logical properties are dependent on what he names linguistic intuitions. These intuitions are not much different from moral intuitions as they are one's sense of right and wrong regarding the conventions surrounding value words. Grenholm finds this aspect of Hare's theory problematic and conceives instead of such intuitions, experiences of having to face moral dilemmas, moral reproach, and praise, experiencing loss, and other similar things a matter of *moral experience*.[85] One may expect one's judgements to be a product of the content in this category. Habermas's analysis is different in this regard and argues instead that the correct logical function of the

---

[83] Richard Hare, *Moral Thinking*, p. 10, 12.
[84] Richard Hare, op.cit., p. 15.
[85] Carl-Henric Grenholm, *Etisk teori*, p. 222, 224, 275f.

word 'right' is located at a higher level and its function is similar to that of truth. As 'it is true that the table is white', he says, 'it is right that one ought not to lie'.[86]

3.2. Critiques of the Concept of Communicative Action

Friedrich Hegel criticised Kant, and the formalistic universality principle of his, of being, in Benhabib's words, '... inconsistent at best and empty at worst'.[87] The criticism, Benhabib explains, was that universalisability does not imply moral rightness; there must also be made some further assumptions about what is good for human beings. Since Hegel, attempts have been made, either to accommodate this criticism, or to creatively try to work around it. Thus, as a consequence, this debate seems to have caused theoreticians to, instead, fall into another trap – that of circularity due to assuming too much. As Benhabib explains, seeming to have Habermas and Karl-Otto Apel, another key figure in the development of discourse ethics, in mind:

> ... either models of practical discourse or the ideal communication community are defined so minimally as to be trivial in their implications or there are more controversial substantive principles guiding their design, and which do not belong among the minimal conditions defining the argumentation situation, in which case they are inconsistent.[88]

Apel, she continues, relies to a larger degree than Habermas does, on what is called a transcendental argument, of which the basic idea is that the sceptic '... is committed to presupposing some basic structural notions ... in order to state his position intelligibly. Because he must in fact presuppose some of the very concepts against which he is arguing, his position is somehow self-defeating'.[89] Apel, Benhabib claims, if I have understood her correctly, is guilty of assuming too much. As Habermas writes, Apel wrongly assumes that the rules of discourse, those necessary presuppositions mentioned previously, governs not only practical discourse, but that they also can be transferred outside of the argumentation situation. However, that one must accept certain rules as a participant in a discourse, does not imply that this rule must be acknowledged outside of it. One cannot derive from the presuppositions of argumentation some 'basic ethical norms', such norms are not the concern of moral philosophy, but instead, the purpose of the philosopher, Habermas claims, is the procedure by which these ethical norms can be justified or rejected.[90]

---

[86] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 53.
[87] Seyla Benhabib, *Situating the Self*, p. 26.
[88] Seyla Benhabib, op.cit, p. 28f.
[89] Michael Power, 'Habermas and Transcendental Arguments', *Philosophy of the Social Sciences*, 23.1 (1993), 26–49., p. 3.
[90] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 85f.

From my explanation in the previous chapter, it should nonetheless be clear that Habermas use this transcendental form of argument too; claiming, like Apel, that a speaker rejecting the rules of discourse commits a performative contradiction. As Habermas does not go as far as Apel, however, the problem for him is that he:

> ... argues that we view [what Apel called the normative constraints of the ideal communication community] as 'universal pragmatic presuppositions' of speech acts corresponding to the know-how of competent 'moral' agents at the postconventional stage.[91]

The problem being that there is no single description of what the know-how of this stage is, suggesting instead that the universal presuppositions of argumentation are contingent, rather than being rules that must necessarily hold. The consequence being, it seems, a similar one that I had identified in the previous chapter where the actor is not rationally bound to accept the universality principle that Habermas has formulated. That is because these rules too, are a product of the individual's particular normative and historical context, or so Benhabib argues.[92]

This argument seems to share certain similarities to White's argument. Indeed, he does not claim that the rules of discourse Habermas has formulated are contingent, but instead, that in the end, the transcendental-pragmatic argument fails because one is not rationally bound to enter the post-conventional stage but may do fine on a less reflective level. That is because such less reflective forms of argumentation still accommodate the obligation to be prepared to justify one's claims to validity.[93] Communicative action corresponds to the post-conventional stage of moral development in presupposing rational commitment to moral principles, rather than speaking of right and wrong in terms of conforming to established norms and roles.[94] The move from the conventional to the post-conventional stage is, as characterised by a move from action being guided by a norm conforming attitude to it being motivated by rational considerations of its validity. When White claims that even the conventional stage presupposes enough reflectivity to satisfy the obligation to provide justification, he must mean that one, from the conventional stage, may still motivate one's actions and statements. Although such motivation cannot reach much farther than that an action is right since it is what is expected.

This analysis also sheds light on an aspect of the problem dealt with in the previous section. Other action theories that presuppose merely one or two actor-world relations, it seems, restricts analysis of moral reasoning to the conventional stage. Theories of rational choice, as it

---

[91] Seyla Benhabib, *Situating the Self*, p. 30.
[92] Seyla Benhabib, op.cit., p. 30
[93] Stephen White, *The Recent Work of Jürgen Habermas*, p. 55.
[94] Jürgen Habermas, *Moral Consciousness and Communicative Action*, p. 125f.

has been explained, view action as motivated by the actor's intentions, of some goal they wish to achieve and similarly reduce language to a medium of strategic influence. 'Role theory', as developed by Émile Durkheim, and later, by Talcott Parsons, view action as norm conforming and language merely as a means to express these cultural values. This explains rather well why Habermas considers these analyses of action and speech to be inadequate, especially considering morality, and gives another dimension to the claim that Hare takes an objectivating attitude to social interaction. Theories that does not allow more than one or two world relations does not provide the tools necessary to grasp all dimensions of action. Thus, applying such perspectives, one will not be able to take actors seriously but render any claims to validity that exceeds one's theoretical framework as illusive or self-deceptive.[95]

I have claimed that this distinction, between a performative attitude and an objectivating one seems reasonable, and to the degree that this distinction corresponds to that of communicative and strategic action, White agrees too, although with reservations against the claim that the former is functionally prior to that of the latter.[96] While White does not seem to think that Habermas need to establish this relationship, I am not so sure. I agree with White that Habermas is having trouble sustaining the necessity of communicative action, but the problem with White's solution is that unless that relation is established, that is, that communicative action is prior to strategic action, Habermas cannot refute moral scepticism. Since his argument depends on that the actor must necessarily accept the presuppositions of argumentation it hinges on the priority of the communicative mode.

As White states, Habermas argues that the adherents to an '... ethical system which keeps 'a dogmatized core of basic convictions' away from the demand of justification' are unable to meet with the 'speech-act immanent obligation' to provide a justification.[97] White does not agree with this, suggesting that the irrationality of the sceptic refusing to act communicatively cannot be applied to someone refusing to '... open up his basic convictions to fundamental criticism'.[98] I do not believe that White's argument is successful however. For instance, someone who makes a rightness-claim based on a conformist attitude to certain cultural values may indeed be willing to justify this claim referring to the general acceptance of those values. The question is whether the discourse will be able to go on much further than this. I do not see how White reaches this conclusion. It does not matter what claims are raised, a rational actor

---

[95] Jürgen Habermas, *The Theory of Communicative Action*, p. 117f.
[96] Stephen White, *The Recent Works of Jürgen Habermas*, p. 45f.
[97] Stephen White, op.cit., p. 54.
[98] Stephen White, op.cit., p. 54.

must be *willing* to provide fallible reasons in favour of the validity of the claim that they have made. They do not have to succeed in persuading the other participants, nor need the argument they give be one which could persuade the participants under certain ideal circumstances. As has been shown, Habermas does not raise such high demands for rationality. Nonetheless, in *refusing* to provide a justification, they will be vulnerable to the same criticism as the moral sceptic since they, at that point, has stopped acting communicatively. Thus, I do not consider White's argument successful, and for the last chapter I will only consider that provided by Benhabib and Dallmayr.

Another attempt to criticise the notion of communicative action is made by Dallmayr. He claims that if one scratch the surface of this distinction, one will see that communicative action too, is oriented towards a certain goal, namely that of agreement. Normally, Habermas seems to want to speak of the communicative mode as a background process, a 'matrix', or a 'system of reference' underlying interaction. Although Dallmayr claims that Habermas is not entirely consistent in this view but does actually speak of consensus and understanding as a goal towards which actors are striving.[99] Dallmayr mentions a number of occasions through the two-volume study where this ambiguity shines through. A particularly distinct example that points towards the latter understanding is a passage where Dallmayr let Habermas explain that communicative action is '... a mode in which "participants pursue their plans consensually on the basis of a common situation definition"', and then adds '... in the same paragraph, "if a common situation definition must first be negotiated ... then consensus ... is itself transformed into a goal or objective"'.[100] Continuing to give similar examples of Habermas saying that all forms of action is teleological,[101] he later concludes that Habermas:

> ... is unable to isolate the communicative category from purposive intent. This is evident already in the adopted terminology: for example, in the opposition between 'success-orientation' and 'consensus-orientation' – where orientation seems readily interchangeable with intention...[102]

This is not what White proposes to be the root of his problem with Habermas. In fact, he tries to refute Dallmayr's argument. As he explains, Dallmayr thinks that Habermas is '... too tied to a subject-centred philosophical orientation. For language is thoroughly instrumentalized for making individual claims and pursuing individual goals'.[103] This concern is not one of mine,

---

[99] Fred Dallmayr, *Polis and Praxis*, p. 237.
[100] Fred Dallmayr, op.cit., p. 238.
[101] Fred Dallmayr, op.cit., p. 240.
[102] Fred Dallmayr, op.cit., p. 240.
[103] Stephen White, *The Recent Works of Jürgen Habermas*, p. 46.

and I will therefore not go into the details of the argument to any further extent than what is needed for me to make my own. There is one problem however. It seems as if Dallmayr should be understood as arguing that communicative action is teleological *in the same sense* as an actor is oriented towards reaching their individual goals.

If so, Dallmayr must be claiming that there is but a minimal difference, if any at all, between the goal-orientation of actors pursuing their individual goal, and the consensus or success-orientations of communicative and strategical action respectively. This is best illustrated by him saying that Habermas is 'oscillating' between speaking of communicative action as sometimes a goal and sometimes a background process. Such an argument seems paradoxical to me. As has been explained, Habermas distinguishes between three types of action: goal-oriented action, where the actor claims to be successfully intervening in the objective world; normatively regulated action, according to which actors claim to be properly following established social norms; subjective self-representations, which means that actors reveals a part of their subjectivity, and that while goal-oriented action raises a claim of successful intervention in the objective world, it is not necessarily strategical. Actors can relate their statements to and act within all three worlds within both modes of action. As Habermas expresses himself, 'In communicative action participants are not primarily oriented towards their own individual successes; they pursue their individual goals under the condition that they can harmonize their plans of action on the basis of common situation definitions'.[104] In the strategic mode then, the opposite holds true, they interact with each other in ways which does not take the others' plans of action into consideration, but *only* their individual success.

Thus, when Dallmayr claims that Habermas is 'oscillating' between speaking of communicative action as a background process and as a goal; he must be suggesting that Habermas sometimes speaks of communicative action implying that actors merely have consensus as the individual goal they are pursuing. Such a claim makes no sense. To my understanding, as is well illustrated in the above quote, communicative action, is a *mode* of action that implies a *consensual attitude* – which is what is meant with background process, I assume. If there is an oscillation, that must mean that sometimes, communicative action moves into the foreground, making consensus an individual goal *instead* of a mode of action. The problem being that this seems to imply that they no longer have the necessary consensus-oriented attitude, meaning that they are strategically oriented or that they act within a hereto unnamed and unmentioned action mode. That is because if they were still having their

---

[104] Jürgen Habermas, *The Theory of Communicative Action*, p. 286.

consensus-oriented attitude, there would be no oscillation. Whatever their individual goals, they act communicatively if they try to, as Habermas says, 'harmonise' their plans with those of other actors.

One can effectively object to this understanding of Habermas, considering his own explanation of the concept. As he writes at one point:

> To avoid misunderstanding I would like to repeat that the communicative model of action does not equate action with communication. Language is a medium of communication that serves understanding, whereas actors, in coming to an understanding with one another so as to coordinate their actions, pursue their particular aims. In this respect the teleological structure is fundamental to all action.[105]

Indeed, while White does not seem to see this as a vindication of Habermas but claims instead that it affirms Dallmayr's point,[106] towards the end of the same paragraph, Habermas concludes his argument in the following way:

> In the case of communicative action the interpretative accomplishments on which cooperative processes of interpretations are based represent the mechanism for *coordinating* action; communicative action is *not exhausted* by the act of reaching understanding in an interpretative manner ... communicative action designates a type of interaction that is *coordinated through* speech acts and does not *coincide with* them.[107]

This last quote is important, as it provides much reason to rethink Dallmayr's critique. If teleological action *is exhausted*, that is, I assume, that the action is 'finished' when it has succeeded or failed, it cannot be a *mode* of interaction. As a mode of interaction, communicative action cannot stop as soon as a participant in interaction has understood the speaker's claim. It seems to be possible to extend this analysis, in the sense that communicative action does not 'end', so to speak, not only when one participant has taken a position on a claim to validity, but also after an agreement has been established. If one is to understand communicative action properly as a mode of interaction, then this understanding appears, at least to me, to be the better one. If that is the case, there is an important difference between communicative action and goal-oriented action. In this sense, what is meant with strategical action, is not simply that it is goal-oriented, but that it, as another mode of action, specifies conditions under which individuals pursues their goals. Rather than the freely entered agreements which communicative action is oriented towards, the strategic mode assumes *only* individual success.

---

[105] Jürgen Habermas, *The Theory of Communicative Action*, p. 101.
[106] Stephen White, *The Recent Work of Jürgen Habermas*, p. 46.
[107] Jürgen Habermas, *The Theory of Communicative Action*, p. 101.

Another question is whether Dallmayr's argument fits with the one given by Benhabib. I have said communicative action in Habermas's theory corresponds to the post-conventional stage of moral reasoning. Considering firstly Benhabib's critique, her argument is that there is no one description of the 'know-how' of this level, but that '... reversibility, universalizability and impartiality, under some description, are all aspects of the moral point of view...' but that the description of these aspects varies.[108] One may consider this argument from two perspectives. According to one of them, one would claim that the fact that differences among philosophers' characterisations of the content of the post-conventional stage is only a problem if Habermas's reasoning entails several interpretations of this content, or the description of each aspect which it is thought to contain. To the degree that Dallmayr's critique is correct, the problem that he has identified, might be what allows several interpretations of the post-conventional stage. This perspective, however, forgets the transcendental structure of Habermas's argument; that is, that he starts from the idea of communicative action, which assumes a particular understanding of this stage of moral reasoning, and infers from that assumption the universality principle. So, from this other perspective which takes the transcendental structure into account, if there is a connection between Benhabib's and Dallmayr's argument insofar as the ambiguous status of communicative action is due to Habermas's assumptions about the content of the post-conventional state of moral reasoning.

3.3. Summary

In the first section of this chapter, I have focused on Hare and Grenholm and their respective versions of prescriptivism. The question being what arguments can be given in favour of, and against, the possibility of rationally justifying moral statements from the viewpoint of prescriptivism. I hold that prescriptivism is able to rationally justify moral judgements.

There are several reasons for this, the main reason being the similarity between Habermas's and Hare's versions of the principle bet there is also the similarity between moral judgements in Hare's theory and Habermas's sincerity claims. Indeed, there are important differences. That Hare does not make use of the different worlds is one such difference, and I doubt that the theory, as he presents it even allows such a distinction. If it does, then it is only the objective and subjective world that can be assumed. Considering Grenholm's modifications something similar to the social world is distinguishable however. Introducing the category of moral experience brings prescriptivism a bit closer to Habermas's position in this sense.

---

[108] Seyla Benhabib, *Situating the Self*, p. 30.

Another important aspect highlighted in this section is a difference in terminology. Speaking about moral principles, Hare seems to refer to what Habermas calls norms, and the universality principle that Hare proposes is thus not a moral principle, but a logical one. Now, this is not something that Habermas would object against as Hare's principle, compared to Habermas's version of it, corresponds to the logical-semantic level of the rules of discourse which Habermas lists.

I also engage with some problems with the concept of communicative action. I consider three philosophers in this section. White, whom identified the problem in the first case, is one of those three, claiming that the distinction is difficult to sustain since one is not rationally bound to enter the post-conventional stage of moral reasoning, and therefore, neither is one bound to accept the universality principle. Benhabib then, claims that Habermas cannot assume that his description of the post-conventional stage of moral reasoning is the only one allowed. If he cannot do that, nor can he assume that his version of the universality principle is the only one allowed since it is entailed by his description of the post-conventional stage. The result in both cases is that the distinction between communicative and strategic action cannot be sustained. This is also what Dallmayr argues, although from a slightly different point of view. His argument is that Habermas cannot describe communicative action only as a mode of action but must always revert to the concept of goal-oriented action. I accept Benhabib's argument, reject White's, and abstain from judgement with regards to Dallmayr at this point, although I find his argument unlikely to succeed.

4. Defining Cognitivism

Now, it seems warranted for me to return to the problem of defining moral cognitivism, which I have not attended to since the introduction. I hope that the progression towards this chapter is as smooth as I intend it to be. Similarly, I hope that my reasons for putting this chapter towards the end of the thesis, rather than at the start which would be the usual place for a chapter having the function that this one has, will be accepted. This section has as one of its purposes to suggest a definition that improves upon those already proposed (4.1.). Besides that, I will also show that Habermas's theory can be considered a cognitivist using this understanding of cognitivism (4.2.).

4.1. Elaborations on the Defining Traits of Cognitivism

As mentioned in the introduction, van Roojen defines cognitivism as (1) moral judgements being truth-apt, and (2) expressing a mental state like that of belief. Accepting van Roojen's definition, I have suggested that (1) touches upon the definition provided by Grenholm since it states that moral judgements are either true or false; and that (2) touches upon that provided by Benhabib since trying to justify one's judgement seems irrational unless one takes moral judgement to be either correct or incorrect, or better or worse. As stated by Habermas in section 2.2., critique of cognitivism is partly due to philosophers being unable to explain what it means for moral judgements to be true. This will indeed be a challenge for anyone arguing in favour of cognitivism to explain. For me, the issue is another. As I do not intend to provide a cognitivist theory or at all argue in favour of cognitivism, I do not see that I would have to provide a solution. However, it seems necessary that in defining cognitivism, in terms of (1) and (2), I will also give some indications as to what they mean. This I think, commits me to at least suggest some restrictions regarding what these consequences these criteria can have regarding plausible cognitivist theories. I must thus explain what it takes for a moral judgement to be (1) truth-apt, and (2) to express a mental state like that of belief.

With regards to (1) then, I will consider moral judgements as being truth-apt if they meet two conditions: (a) that they have assertoric content, and (b) that they have correctness conditions. The former means that they have the linguistic form of propositions and thus certain logical properties. In brief, as Lillehammer suggests, they may feature in premises in logically valid, that is, truth-preserving inferences.[109] As for (b), if cognitivism is correct, this criterion specifies what it would mean for moral judgements to be true. While I follow Lillehammer

---

[109] Hallvard Lillehammer, 'Moral Cognitivism', *Philosophical Papers*, 31.1 (2002), 1–25., p. 5.

rather than Grenholm in so doing, I do not agree with either on what it takes for this criterion to be met. Most notably, my understanding is distinctively weaker than Grenholm's, while at the same time, it avoids the deflationism favoured by Lillehammer. My reason for rejecting both accounts is that I do not want to discriminate between substantial theories about truth, either by assuming that truth means, say, correspondence to facts, but neither do I want to deny that truth is a substantial property as deflationists do.

As such, I do not consider it necessary to specify whether truth means correspondence to facts or coherence with other true statements, or any other substantial understanding. Nor do I take truth-aptness to imply either realism or anti-realism concerning values. Instead I assume only a minimal definition of truth as meeting the so-called t-schema. If my understanding is correct, a substantial truth-definition must meet the t-schema for it to be considered plausible.[110] Originally developed for formalised languages, it states that: '"it is true that *p*", if and only if *p*', and as adapted to natural languages by Davidson, it is reformulated into: 'sentence *s* is true (as English) for speaker *u* at time *t* if and only if *p*', for as he explains, the simple formula that is the t-schema '... cannot be ours; for when there are indexical terms (demonstratives, tenses), what goes for "*p*" cannot in general be what "*s*" names or a translation of it...'.[111]

What is being said is two things: (i) that the metalinguistic expression 'it is true that *p*' is logically equivalent to simply stating *p*; (ii) for a proposition to be true there are certain conditions which need to be satisfied. (i) has often been interpreted as meaning that the truth predicate is redundant since one makes a truth-claim merely by saying *p*. Proposing that 'snow is white', one says something about the world, while, in saying that 'it is true that snow is white', one says instead something *about the proposition* that snow is white, hence the distinction between object language and metalanguage. As for (ii), this insight is taken farther and presupposes a relation between language and world which is called *satisfaction*. Take the example 'snow is white', one can see that what must be satisfied for the statement to be true is that snow is white. This example is fairly simple statement (although Davidson expresses some objections,[112] I use it as an illustrative example) and matters are easily complicated. Take instead the statement 'it is Tuesday'. For this to be true in terms of satisfaction it must be Tuesday in the speaker's language and time zone at the time the statement is made.[113]

---

[110] Tomas Ekenberg and others, 'Språk, Värld, Sanning', in *Kunskapsteori: En Introduktion till Vetande, Berättigande Och Sanning* (Lund: Studentliteratur, 2014), pp. 91–124., p. 118.

[111] Donald Davidson, 'True to the Facts', in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 37–59., p. 45, 46.

[112] Donald Davidson, op.cit., p. 50.

[113] Donald Davidson, op.cit., p. 45, 47f.

I do not claim to provide a substantial truth-definition here. It is a minimal definition of truth in the sense that it only specifies some criteria for truth-aptness, that is some basic standards that must be met for a sentence to be able to make a claim to truth. In this sense, whether Habermas is a cognitivist must be judged according to his own definition of truth, that is the case for Grenholm and Hare too. Now, this does seem to me to imply that Grenholm does not accept cognitivism according to my definition of the concept. As he explains:

> Moraliska omdömen saknar sanningsvärde eftersom de inte uttrycker påståenden om något sakförhållande. De konstaterar att en erfarenhet är god eller att en handling är rätt, men deras primära funktion är en annan. Den är att uttrycka uppmaningar eller befallningar…

> [This roughly translates into the following: Moral judgements are not truth-apt since they do not express propositions about states of affairs. They may propose that an experience is good or that an action is right, but their primary function is another. It is to suggest courses of action or to issue commands...].[114]

This objection is not so much directed to moral judgements being propositions, it seems to me, but is rather directed at their content. It does not follow from the linguistic form of moral judgements that they are actually used to state that something is the case.[115] Hare's idea is similar. As previously shown, he claims that anyone speaking about truth and moral judgements must be confusing moral philosophy with some empirical science like anthropology. This objection is most strongly directed towards intuitionist theories claiming that one's access to moral knowledge comes from one's moral intuitions. Rather than providing a source of moral knowledge, it is from moral intuitions that relativism gets its strength. It is not by any means necessary for some person's intuitions to be the same as someone else's, and if intuitions are the sole source for moral knowledge, one must give all intuitions equal consideration.[116]

It should be acknowledged before moving on that Grenholm seems to allow metaethical theories that defines truth in terms of coherence, and that speak of possibilities of having practical knowledge about the rightness of moral judgements to be cognitivist, although he, as seen in the above quote and his initial definition of the concept, seems to reject such theories.[117] A coherence theory of truth does not align with this definition of cognitivism since a true statement, according to such understandings of truth, is not one which relates to some obtaining

---

[114] Carl-Henric Grenholm, *Etisk teori*, p. 268.

[115] Hallvard Lillehammer, 'Moral Cognitivism', p. 8. Lillehammer notes at this point that several non-cognitivists have made such an argument.

[116] Richard Hare, *Moral Thinking*, p. 12.

[117] Grenholm refers to Christine Korsgaard, John Rawls, and Carla Bagnoli as Kantian constructivists, whose theories are cognitivist although their value ontology is a negative one. See Carl-Henric Grenholm, *Etisk Teori*, p. 260.

state of affairs but to other true statements.[118] The consequence being either that Grenholm's focus is on denying that these philosophers are cognitivists, but rather to argue in favour of his own conception of cognitivism, and that this gives rise to this implicit rejection of their understanding of the concept. It might also be the case that I am understanding Grenholm wrong. If by true sentences expressing states of affairs which obtain, he means something similar to how I have defined truth-aptness, so that a positive ontology is not required, then my previous interpretation seems unfair. However, if that is the case, the above quote seems odd since moral statements would be truth-apt.

Considering (2) instead, in uttering a moral judgement, the speaker is taken to express a mental state like that of belief, or more precisely, they express a belief-like *propositional attitude*. Such an attitude can, rather bluntly, be said to be an attitude towards a proposition. Consider the following example:

> A little boy *wishes* that he had his favorite toy horse, *sees* that it is on top of his dresser, and *decides* that he will fetch it. *Believing* that he is supposed to stay in bed he slips out quietly, *hoping* that he will not be caught but *fearing* that he might [emphasis added].[119]

As is then noted, wishing, perceiving, deciding, believing, hoping, and fearing are attitudes towards the different propositions following 'that': that the toy is on top of the dresser; that he will fetch it; and that he is supposed to stay in bed,[120] to mention a few. As can be seen above, in believing a proposition, it is taken to have correctness conditions: the boy believing that he is supposed stay in bed, for instance, could mean that he believes that his parents consider it too late for him to be awake at this time at night. The boy thinks that a state of affairs obtain. I take this criterion to state, not that moral judgement *has* correctness conditions or is true or false in any sense, which is specified by (1), but instead that moral judgements are *taken* to have correctness-conditions.

Lillehammer speaks of this in a clarifying, although slightly different way, calling it *world-guidedness*.[121] In being world-guided – not excluding them from being *world-guiding* at the same time – a central function of moral judgements is to reflect the world. Considering the above quote, it becomes obvious that world-guided propositions can express several attitudes taken by the actor towards the proposition. Fearing and hoping seems to be two examples. These propositional attitudes might spring from uncertainty about whether or not the boy's parents are

---

[118] Carl-Henric Grenholm, *Etisk teori*, p. 262.
[119] Timothy Schroeder, 'Propositional Attitudes', *Philosophy Compass*, 1.1 (2006), 35–73., p. 1.
[120] Timothy Schroeder, "Propositional Attitudes", p. 1.
[121] Hallvard Lillehammer, "Moral Cognitivism", p. 11.

still awake and whether or not they might hear him. But there are other forms of world-guided attitudes too. The boy might *imagine* that he has his toy horse to give an example. Lillehammer correspondingly distinguishes between attitudes which are *permissively* world-guided and those which are *non-permissively* world-guided. Hopes, fears, and imaginings are of the former variety, whereas beliefs are of the latter.[122] In imagining that he has his toy horse, the boy might be picturing it in his mind; imagining how it looks and what it feels like, or maybe how he would be playing with it. Fearing then, in this particular case, involves both believing that he ought not get out of bed, and that there is at least some possibility of him getting caught trying to fetch his toy. It is the same with regards to him hoping that he will not get caught. Permissively world-guided propositions do not have correctness conditions, whereas non-permissively world-guided one's have. One might ask why this distinction is necessary. It seems to me however, as has been said before, that when arguing about moral judgements, if a *judgement* is to have any form of authority distinct from the speaker uttering it or any relations of power between the speaker and the hearer; that is, if the judgement to be able to rationally motivate the hearer, it needs to be non-permissively world-guided. If it is not, then it can be dismissed as merely imaginings or wishful thinking.

Understood in this way, it is not obvious that (2) is insufficient on its own. The criterion does specify the rather crucial aspect that moral judgements express a cognitive mental state, that of belief to be precise. It also specifies that a central function of moral judgements is to correctly represent the world to which they relate – that they are non-permissively world-guided, that is. Considering Lillehammer's understanding of world-guidedness, it would also include the correctness conditions for such judgements. For Lillehammer, the reason is that without his account of truth-aptness, that moral judgements take the linguistic forms of propositions, or in his words, that they have assertoric content, it is by no means established that moral judgements are cognitive. For instance, a thermostat is world-guided while any results of its readings cannot be said to express a mental state like that of belief. This is not a concern of mine however, since thermostats does not have propositional attitudes or any other form of mental states. However, taking a proposition to have correctness conditions does not mean that it does, therefore, (1) is necessary; similarly, if no one takes a proposition to have correctness conditions it does not matter much whether it does so or not, which makes (2) necessary too.

---

[122] Hallvard Lillehammer, 'Moral Cognitivism', p. 11, 12.

4.2. Truth-aptness: Moral Realism or Validity Conditions?

For anyone having a similar understanding of cognitivism as Grenholm and Hare, it might seem problematic to call Habermas a cognitivist. Even following the definition of cognitivism that I have proposed, as the view that moral propositions are (1) truth-apt, and (2) that they express a mental state like that of belief, it is not obvious whether or not truth-aptness can be applied to rightness. It seems uncontroversial to say that the discourse theory of morality satisfies (2). In acting communicatively, a person seeks to represent whatever world they relate their statement to correctly. It also seems as if each of the three worlds *can* be represented correctly or incorrectly according to Habermas's theory. What might be more controversial is to propose that rightness-claims are truth-apt. Finlayson claims that Habermas does not think that this is the case. Doing this, he relies on a similar understanding of cognitivism as Grenholm does, defining cognitivism only in terms of their truth-aptness understood in terms of moral realism. In order to avoid the ironical conclusion that Habermas, while claiming that cognitivism is a necessary condition for justification of moral judgements, is not himself a cognitivist, Finlayson distinguishes between *metaethical* cognitivism and *moral* cognitivism and suggests that Habermas only rejects the former.[123]

His position relies on three claims: firstly, that Habermas holds moral rightness and truth to be *only* analogous and that the similarities between the concepts are not enough to consider moral judgements truth-apt; secondly, that Habermas, by rejecting theories about moral realism rejects metaethical cognitivism too; thirdly, that Habermas's theory is susceptible to the Frege-Geach problem.[124] I will not debate these issues with Finlayson to any further extent than showing why the concept of rightness satisfies the criterion of truth-aptness, and why Finlayson is wrong in claiming that Habermas rejects metaethical cognitivism. I will not consider the Frege-Geach problem since the outcome of such a discussion does not affect whether or not Habermas is or is not a cognitivist or what cognitivism means, only whether his cognitivist or non-cognitivist position is or is not tenable. While it is surely an interesting discussion, it is beyond the scope of this thesis.

What I must do, I believe, is firstly to mention the reasons that I have presented in previous chapters for rejecting Finlayson's claim that there are not enough similarities between truth and rightness to accept rightness as truth-apt; secondly, I must show that Habermas's conception of truth satisfies the t-schema and that rightness-claims does this too. If my

---

[123] James Gordon Finlayson, 'Habermas's Moral Cognitivism and the Frege-Geach Challenge', *European Journal of Philosophy*, 13.3 (2005), 319–44., p. 320.
[124] James Gordon Finlayson, op.cit., p.320f.

argument fails at these points, then Habermas cannot be considered a cognitivist as I have explained it. If, however, both conditions are met, I will claim that I am justified in considering the overarching concept of validity as corresponding to the criterion of truth-aptness and consequently, that Habermas's theory satisfies both criteria explicated for moral cognitivism. So, firstly then, there are much evidence in Habermas's texts for removing the word 'only', and thereby acknowledge that truth and rightness simply are analogous (as opposed to *only* analogous). Recall for instance Habermas saying that feelings have a similar relation to the social world as one's sense perceptions has to the objective one (see section 2.1.), and that the universality principle acts as a bridging principle analogous to the principle of induction (see 2.2.). Consider further, that for the objective and the social world alike, the participants have no ontological assumptions about the existence of an objective world, but that both worlds gain objectivity, if still in different senses, as the participants starts to consider it as one and the same. The social and objective world alike becomes the same through the process of moral-practical discourse and theoretical discourse respectively as participants reach rationally motivated consensus (see 2.2.). Indeed, there are differences between the two worlds too: that the objective world *is* a world independent from the human mind is one such difference, and the action-motivating nature of social norms contrary the purely factual one of objects in the objective world is another important difference. But that is what warrants making a distinction between truth and rightness, not treat them as distinct, although in some respects overlapping concepts. Both concepts are forms of validity and are different insofar as the world to which they relate differs from the others.

Finlayson's argument fails for another reason as well. As he says:

> [Habermas's] main objection to metaethical cognitivism is that to grant truth to some moral statements is to raise the spectre of moral realism. Moral realism, as Habermas understand it, involves the positing of mind-independent (or language-independent) moral values and properties, and thus threatens to burden moral theory with an outlandish ontology.[125]

Here, Finlayson only establishes two things: that Habermas rejects an already problematic understanding of cognitivism and that he rejects moral realism. That one must distinguish truth from rightness has to do with Habermas relating truth-claims to the objective world. This is exemplified in the beginning of section 2.2., where he is quoted saying that sceptical critique of cognitivism is, among other things, directed towards philosophers being unable to explain 'what it would mean for moral propositions to be true'. The distinction between truth and

---

[125] James Gordon Finlayson, 'Habermas's Moral Cognitivism and the Frege-Geach Problem', p. 320.

rightness and their integration into the concept of validity are attempts to respond to that critique.

While I believe that Finlayson's argument has been sufficiently defeated, my argument needs further support. I must show that the discourse theory of truth lives up to being truth-apt in the sense explained. While I do believe that such a claim is rather uncontroversial, it seems proper to give some reasons for accepting it. Habermas's truth-conditional semantics is one such important reason. Indeed, this is, to Habermas, a theory about the meaning of sentences and not one about truth, but as Davidson notices at several times: meaning, belief and truth is tightly intertwined. One cannot know what someone means without knowing what they believe, nor can one know what someone believes without knowing what they mean with their utterances, as it is those that contains their beliefs.[126] Davidson's solution is the truth-theory he proposes. Claiming that a sentence is true, the speaker claims that the conditions for the truth of that sentence are satisfied. Knowing those conditions, the hearer will be able to figure out both what the speaker believe and what they mean since a statement, say, 'snow is white', if it is genuinely stated, contains the belief that snow is white, and also, the hearer, who must assume that the speaker is sincere, will know that what the speaker means to say is that snow has the colour white. Being able to figure out the meaning of sentences from their truth-conditions is due to established norms of proper language use.

As Crispin Wright notes about proper assertoric practise: '... unless participants in the practice for the most part try to respect the norms of warranted assertion which govern it, it is not clear in what the fact could consist that its ingredient sentences have the content which they do'.[127] If there were no norms of proper language use, sentences would not have any truth-conditions or correctness conditions of any form. I take that what is being described here is similar to what Habermas speaks of as the coming together of a world as one and the same for the participants in a practical or theoretical discourse (see section 2.2.).

While Habermas does not mention the t-schema explicitly, he seems to accept it as he, when arguing that 'right' is more similar to a truth-predicate than some other lower one, he says that just like claiming that 'the table is white', is logically equivalent to 'it is *true* that the table is white', the same structure can be found in claims to rightness; that 'one ought not to lie', is

---

[126] Donald Davidson, 'Truth and Meaning', in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 17–36., p. 24; Donald Davidson, 'Belief and the Basis of Meaning', in *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), pp. 141–54., p. 144.

[127] Crispin Wright, 'Inflating Deflationism', in *Truth and Objectivity* (Cambridge: Harvard University Press, 1994), pp. 1–33., p. 17. I will not follow Wright's argument any further since it goes in a somewhat different direction than that which I have in mind.

equivalent to 'it is *right* that one ought not to lie' (see section 3.1.). This is good evidence, I would say, that Habermas's theory meets the t-schema, but also that Habermas argues that rightness has correctness conditions. Anyone seeking the conditions for the truth of a proposition will, I believe, accept the answer that the t-schema provides. Objections might be raised against it being truistic, but further explanations and adding the nuances that Davidson gives it might, I believe, remove such dissatisfactions. However, applied to rightness this is not equally intuitive. Take the sentence 'one ought not to lie', fed into the t-schema, one might give it the following structure '"it is right that one ought not to lie", if and only if one ought not to lie'. This sentence does not appear to be, at least to me, nowhere near as intuitively correct. It is not obvious what conditions are referred to by that statement; what conditions are to be satisfied by the world to which it relates. That is not so much a problem of the theory itself, since the comparison of sentences might be unfair in the first place. As showed in the previous section, while the statement that 'snow is white' is rather simple, the statement that 'it is Tuesday' immediately becomes more complex, and there are certainly those statements that has an even more complex structure. It might be the case that 'one ought not to lie' is a statement more fairly compared with these latter, more complex, statements. I shall not attempt to provide an equivalent of a t-sentence for rightness claims in the same sense as Davidson has done for truth. I will settle with explaining where one might find these validity conditions.

This was, to some extent, explained in 2.2., where I first gave an account of Habermas's truth-conditional semantics. Here, I shall extend on that analysis somewhat. Firstly, Habermas account of truth conditional semantics differs somewhat from Davidson's even regarding its basic structure. Regarding ordinary truth-claims, in this particular case, a weather prediction, Habermas suggests that a hearer understands the claim if they know '... (a) the conditions that would make the prediction true, and (b) the conditions under which [the speaker] could have convincing reasons for holding a statement with the content (a) to be true'.[128] This difference between them does not affect anything in terms of the compatibility between my understanding of truth-aptness and Habermas's analysis of a truth-claim. The satisfaction of correctness-conditions remains the central aspect, and that, I believe, is the important part. What Habermas's has done is to extend the conditions to be satisfied from merely the truth of the statement, to incorporate also what would be sufficient reason to assume that, in this case, the prediction, will come true. I assume that this difference is due to the merging of truth-conditional semantics and speech-act theory. When accepting or rejecting a claim, the hearer

---

[128] Jürgen Habermas, *The Theory of Communicative Action*, p. 303.

has not only the literal sentence to take a position towards, but also towards what it is that the speaker is doing – in this particular case, predicting the weather. The structure of rightness-claims is the same. Understanding the proposition 'tomorrow will be a sunny day', relies on knowing when these conditions are satisfied, that is, that one knows what it means for a day to be sunny, and to when 'tomorrow' refers. To take a position on the rightness of the claim 'one ought not to lie' one needs to know such things as whom is affected by this norm; one must know what a lie is, and, I will assume, one must know the logical function of 'ought'. But this is only what corresponds to (a) in the above quote, it is also required that the hearer, as is specified by (b), knows what it would mean to have sufficient reason to consider the claim as valid.

What it is that (b) actually specifies is not entirely clear to me, not regarding the truth-claim nor the rightness-claim. However, I will assume that these conditions, considering first the weather prediction, refers to the prerequisites the speaker actually has to make such a prediction. It might refer to their access to instruments with which they can make such prediction with at least some degree of accuracy, or if they are just referring to their intuition. One might in other words assume that the conditions specified by (b), would be satisfied by the meteorologist making a prediction that 'tomorrow will be a sunny day'. Trusting the meteorologist might also fall under these conditions, whereas 'I would feel it in my bones if a storm was coming', does not. Considering rightness then, these conditions must refer to what was discussed in chapter 2.1., one's moral experience, and the free assent by all affected as participants in a practical discourse.

4.3. Summary

In this chapter, I have elaborated upon my definition of cognitivism, explaining more thoroughly what the two criteria mean: (1) that moral statements are truth-apt, which means that they (i) have the linguistic form of propositions, and (ii) that they have correctness conditions; and (2) that the propositional attitude generally taken towards moral statements is that of beliefs. If cognitivism is true, both criteria must be met.

As for (1), my discussion on (i) was rather brief, that moral statements have the form of propositions means that they share certain logical properties with them; my discussion on (ii) was more thorough, leading to the conclusion that moral proposition, for them to be able to say something which is either true or false, certain conditions must be given which is either satisfied or not. I chose this relation of satisfaction, rather than correspondence to facts, or coherence with other true sentences so that the theory will not discriminate against substantive truth-

theories. This criterion of truth-aptness does not say anything at all about the nature of truth, but only what it takes for a statement to be able to say something which can be either true or false. Regarding (2) then, that moral propositions expresses a mental state like that of belief means that the propositional attitude taken by speakers and actors, if cognitivism is true, are that of beliefs. This is an important criterion since if the actors generally do not take an attitude like that of belief towards moral propositions it does not matter much whether they can be true or false or not.

In section two, the focus is on the notion of truth-aptness and correctness conditions, especially with regards to Habermas's theory and his truth-conditional semantics. I try to simultaneously argue that Habermas is a cognitivist, that Hare and Grenholm remains non-cognitivists, and justify my decision to give a definition of cognitivism rather than a cognitivist metaethical theory.

5. Concluding Remarks

In the following two sections, I will try to show how all these, somewhat diverging threads fit together. I begin in section 5.1. with a discussion on the universality principle and will try to decide whether or not Hare may rationally justify it. In order to do so, there are several things that needs to be sorted out: (1) what the function of the moral principle is, and whether it is necessary that Hare's principle is a moral priniciple; (2) the connection between correctness conditions and rational justifications, and (3) whether Hare can rationally justify his principle. In section 5.2., I consider the problem I identified with the distinctions between communicative and strategic action on the one hand, and the participant's and the observer's attitude to social interaction on the other. To settle this dispute, it must be explained (1) whether there is a connection between Benhabib's and Dallmayr's arguments; (2) what issues the problem with the distinction cause for Habermas's argument, and (3) if the argument against the distinction is successful.

5.1. The Rational Justification of Moral Principles

A first issue which must be sorted out is the role of principles in the respective theories. Habermas for instance, speaks of principles on the one hand, and norms on the other. Among the principles are most notably the universality principle and the discourse principle, the latter of which he calls a moral principle. Norm are something else and is more correctly referred to as patterns of behaviour. Hare does not make this distinction but speaks only of principles. This is indeed rather confusing and is an obstacle in understanding Habermas's critique of Hare. Habermas is correct in concluding that Hare's principle is not a moral one, if one considers a moral principle to be a principle with normative content. From this point of view, Habermas criticises Hare for not including, in his version of the universality principle, a notion of impartiality. It is this lack of impartiality that makes the validity of moral judgements something subjective on Hare's view, since the only condition for making a valid one is logical consistency. This much has already been shown, and apart from one thing, it seems to me to be a fair critique.

This one issue that I have, is due to the confusion with regards to principles and norms. I do not believe that there is so much of a difference between Habermas and Hare on this matter as it might seem to be considering the criticism that Habermas directs against Hare. When Hare says that one cannot give a justification for a way of life, it seems as if he says that whatever judgements someone makes, they will always be the product of a principle, which in its turn is a product of their experiences and web of attitudes and social relations in which they live their

life and similar contingencies. As Hare was quoted on his view of the impossibility of a complete justification (see section 3.1.), the choice of principle he speaks about is not a choice of a principle like Habermas's universality principle that may decide between valid and invalid moral norms or judgements. It seems to me, considering how Hare speaks about "ways of life", that this choice of principle is more of a choice of a highly abstract norm which the actor takes to be valid and in the light of which they decide how to make more concrete judgements. I see no reason that these norms nor the judgements made on the basis of them, would be out of reach from Hare's version of the universality principle.

My point is, that it does not seem to me as if Habermas has any good reason, considering his own theory, to disagree. According to Habermas too, persons live their life within contexts that differs from each other to a higher or a lesser degree. These contexts are characterised by different norm complexes on the basis of which individual actors make their individual judgements. It is due to this, that Habermas formulates his principle the way that he does, in order for it not to favour certain normative contexts over others (see for instance, his critique of Kant's categorical imperative which was briefly mentioned in 2.1.). Nonetheless, if some norm is rendered problematic, all affected engage in practical discourse attempting thereby to determine its validity by rational argumentation. The reasons put forward in defence of a claim or to reject it will be based on the speaker's particular point of view, as is the case according to Hare too. When all freely accepts the outcome, the debated norm has either been rejected or once again accepted as valid, this outcome might however differ between different contexts due to the contingencies mentioned earlier.

This raises questions about the connection between cognitivism and rational justification. One problem is that it seems that for one to be able to reason rationally about something, there need to be some form of correctness conditions; making a rational argument the conclusion needs to, if not follow deductively then at least it needs to be supported, inductively for instance, by the reasons adduced in favour of the claim. But if Hare and Grenholm are non-cognitivists, which they claim to be, and as I have suggested that they are in my discussion on cognitivism in the previous chapter, it does not seem like they can assume that moral judgements have correctness conditions or considered better or worse as it is a characteristic of cognitivism.

As has been seen, Habermas's position is that given ideal circumstances, practical discourses will end in consensus; if the rules of discourse which were mentioned in section 2.1. are acknowledged by the participant's, and consequently, that they are genuinely oriented towards reaching an understanding, they will recognise and accept the better argument. This recognition cannot be based on the contingencies of a particular context since that would ruin

the argument Habermas is trying to make. While different reasons might be provided in different contexts which also might produce different outcomes, the correctness conditions are not. Hare's position is very different on this matter. As I showed above, he seems to hold that agents must assume, considering their experiences of interaction with other human beings what interests and inclinations humans generally have, and from these assumptions then figure out how to behave and live. These conclusions are not subject to being right or wrong in any sense, but unless they are consistent with one's own inclinations, one's sincerity, or rationality is doubted. Thus, one may reach the conclusion that moral judgements have correctness conditions according to Hare too. There certainly are correct and incorrect moral reasoning, and therefore, one might claim, that there will be conclusions that are correct or incorrect. However, one's judgement can only be correct relative to one's own premises. I previously compared the status of moral judgements in Hare's theory with Habermas's claims to sincerity, and I think that this comparison can at least explain the form of rational justification that Hare's theory allows. If one can talk about correctness or validity in relation to Hare, it is merely a claim that one is honest about what one says about one's own views, a claim that may be evaluated only with regard to the speaker's behaviour.

One might wonder whether this should not be considered a form of cognitivism. A sincerity claim does indeed meet the t-schema, it does also, clearly, have validity conditions. It becomes troublesome however, considering the second criterion that it expresses a mental state like that of belief. If a speaker states that 'one ought not to lie', to the degree that this proposition expresses a rightness-claim, it expresses something that the speaker believes; however, to the degree that it is a sincerity claim, if cognitivism should be extended to this form of validity, this belief that it is right that one ought not to lie must also include a belief of the speaker about their own sincerity. However, if the speaker is not sincere, then the proposition 'one ought not to lie', would not express a mental state like that of belief. It is not compatible with being a belief that the statement is insincere. Thus, it seems odd to speak about beliefs about one's own sincerity. One might have beliefs about the external worlds (the social and objective) since one might consider the available evidence and draw a conclusion which one believes to be correct. Even if it is wrong, it remains a belief until the holder of the belief has been genuinely convinced that it is, in fact, wrong. One cannot approach one's own sincerity in this way however, one cannot believe to be sincere but be insincere. One can, as Grenholm points out to Hare, make a judgement, be sincere about it but still fail to act in accordance with it; one can also, it seems to me, believe to have certain, say, reasons for having performed an action, and be wrong about it, but that does not render Hare a cognitivist.

There is also another point that needs to be made at this time. When White states (see the end of section 2.2.) that a non-cognitivist, in contrast to Habermas, would not consider the universality principle justifiable, it seems to me that he misses the difference in terminology that I noted in the beginning of this section. At least if one, when talking of non-cognitivists, refers to universal prescriptivism as developed by Hare. The decision of principle that Hare speaks about (see section 3.1.) is *not* a principle like Habermas's universality principle. As Hare speaks of moral principles, they do not merely specify how to behave against others but also includes an idea about the good life. Habermas's universality principle specifies neither, it places some restrictions on which ideas of the good life may be pursued, but so does Hare's principle, although to a lesser degree and only relative to the individual's subjective view. This is important, I believe, since it seems to me to be a source of misunderstanding. Habermas's critique of Hare, and White's analysis of it, relies, at least partly, on what might be boiled down to a terminological confusion: what they mean with the word principle, and especially moral principle.

Indeed, Habermas's critique of Hare considering moral judgements to be a form of imperatives still has weight. He is right I believe in that if Hare's analysis is correct, argumentation is not *required*. Participants in a discourse may argue all they want, no one will be able to claim to be right since agreement is not and cannot be rational due to the force of an imperative being located in whatever sanctions or rewards the hearer believes that the speaker can make effective rather than the argument being a good one. I do think that this is a misrepresentation of Hare's theory however. As has been shown (section 3.1.), Hare does seem to agree with Habermas on the inappropriateness of using such strategies to force an end to the argument. That speaks in favour of recognising that Hare's theory might be prone to have even more rational potential than even my own analysis has allowed, since Hare, from this point of view, seems to allow for some amount of intersubjectivity.

As seen, Grenholm modifies Hare's theory in some respects. The most important one being that he considers the considers the category he calls "moral experience" rather than merely the idea of linguistic intuitions as relevant for making reasonable moral judgements. Considering morality as a social institution,[129] it seems to me as if he gives some further sense of a common ground from which persons may reason. Indeed, this does not mean that persons have similar experiences of this institution. One might, it seems to me, be able to draw a parallel

---

[129] Carl-Henric Grenholm, *Etisk teori*, p. 260. Grenholm states here that morality, like money, and universities are 'institutional facts' that are created and further developed by human interaction.

between this category, and Habermas's social world. I am not able to do a detailed analysis of the similarities, but he too seems to picture norms as belonging to a particular social domain.

I do not think that this changes much however. That a norm exists, or even, that it is generally accepted does not make it valid, and for Grenholm, if one can speak of valid norms, it is not validity in the same sense as Habermas uses the word, but one which meets a universality principle like that proposed by Hare. Grenholm also presupposes that one might evaluate moral statements by their coherency, whether they fit with one's commitments, knowledge about oneself and the reality, and as previously mentioned, one's moral experience.[130] None of these criteria can however carry his view of moral statements over to the other side of the 'cognitive/non-cognitive divide', to borrow van Roojen's words from the introduction. As each criterion is formulated seems to mean that one's own moral judgements must be, say, coherent, with each other, and not with someone else's judgements.

Furthermore, considering the understanding of truth, it seems that if moral statements would be truth-apt, to Grenholm as well as to Hare, that would imply moral realism. One might want to consider the differences in their concept of truth and truth-aptness. One of the reasons I have for arguing in favour of a minimalist definition of truth, as previously mentioned, is that I do not think that a definition of cognitivism is the place to provide a full-blown theory of truth, but only to explain what it takes for something to raise claims to truth, that is, to be truth-apt. This means that one ought to combine this notion of truth-aptness with a more substantial theory of truth which gives an account not only of the epistemological and semantic level, but also of the ontological one. Considering Grenholm's, as well as Hare's, understanding of truth as expressing propositions about some state of affairs, one must consider him a non-cognitivist since this conception of truth as applied to moral judgements seems to imply a positive value ontology were they to be considered cognitivists. Neither Grenholm nor Hare would accept such a conclusion. That Habermas also considers truth-claims to relate to the objective world is not an issue for considering Habermas a cognitivist since the important concept, with regard to a proposition's truth-aptness, in his theory, is the concept of validity rather than that of truth.

5.2. The Priority of Communicative Action

The line of argument that this side of the thesis has been pulling, began in the insight that it does not seem like Habermas's claim that actors are rationally bound to always act communicatively is correct. The initial problem for me was that it does not seem like there are

---

[130] Carl-Henric Grenholm, *Etisk teori*, p. 274ff.

situations where acting strategically, or at least retract from the participant's perspective, might in fact be necessary. From there, three accounts of the problem has been given: firstly, Benhabib argued that there is no agreement on how the content of the post-conventional stage of moral reasoning should be understood, and therefore, that Habermas is unable to conclusively establish that communicative rationality is the correct one; I then considered White's argument that Habermas's attempt to establish the conceptual priority of communicative action fails since less reflective levels of moral reasoning does meet the criteria of standing ready to justify one's claims, I deemed this unsuccessful; thirdly, and lastly, I considered Dallmayr's claim that the two modes of action is actually not that different at all, but that both relies on the notion of goal-oriented action. I have then said that to the degree that Dallmayr's argument is correct, it might be explained by Benhabib's claim that his description of the post-conventional stage of moral reasoning is not the only one that is plausible. Dallmayr's argument might then, if successful, explain why Habermas seems to have a hard time to establish the conceptual priority of communicative action. The point that am trying to make here, is that if both these arguments are successful, then it becomes difficult for Habermas to sustain the necessary priority of communicative action and reason, over the strategic mode. If this distinction collapses, then the claim that non-cognitivists cannot rationally justify moral judgements, norms, or principles is rendered problematic too.

I must start with determining the success of the two arguments. I have already indicated some problems with Dallmayr's argument, but there are some reasons to consider that one provided by Benhabib as unsuccessful too. The argument that different philosophers disagree on the description of the content of the post-conventional stage of moral reasoning cannot in itself be a problem. One cannot be required to begin one's argument somewhere where there is no or merely some disagreement. Habermas does claim to provide a better description of the post-conventional stage than his opponents. His argument against monological descriptions of universality which was briefly mentioned and described in chapter 2.1., being one example. If his argument is a plausible one, he must be allowed to reject other accounts of this stage of moral reasoning. The fact that there are many *possible* accounts of this stage of moral reasoning is not a good argument against arguing in favour of one particular account.

However, it might also be the case that Benhabib thinks that there are several *reasonable* accounts of the post-conventional stage of moral reasoning, or at least that deciding such things are beyond the scope of moral philosophy. She says that: '... the formal structure of postconventional moral reasoning allows a number of substantive moral interpretations, and these interpretations always take place by presupposing a hermeneutic horizon of norms and

principles which has become aspects of a modern lifeworld'.[131] I take that the 'formal structure' mentioned is the content of the post-conventional stage, and I take the 'substantive moral interpretations' to refer to the way that this content, impartiality among other things, is defined and what principles this description of the post-conventional stage are thought to entail.

While one may fault her for not taking a position on this matter, there are good reason for abstaining from doing so too. From the point of view of discourse ethics, it has the benefit that it is not merely the validity of some norm that is up for discussion but all normative assumptions on which the theory is built. There are also epistemological concerns for not taking a position on the matter which are similar to those which made me abstain from taking a position on the nature of truth in my definition of cognitivism. Simply, that while there might be a correct description of truth, or in this case the post-conventional stage of moral reasoning, there is no consensus on which one it is. Since being justified in believing that one of several plausible accounts is correct, does not ensure the correctness of this particular description, one's theory must remain agnostic. Such a position has problems too since it can render the whole project of theory construction impossible, since for the same reasons, one cannot be sure that uncontroversial or even generally accepted assumptions are true. I will nonetheless accept this as a viable way of dealing with some controversies that might arise when constructing a theory. Whether or not it works for Benhabib with regard to her theory, is a discussion for another time, for the purpose of this thesis I will assume that it does.

What about Dallmayr then? Where this discussion stands at this time, the question is whether or not Dallmayr is correct in his claim that *oriented* towards reaching an understanding, means having the *intention* to reach an understanding. The challenge that this argument faces is that if it is to be a successful one, one must be able to argue that, for Habermas, 'orientation' is interchangeable with 'intention' without, at the same time, arguing that, as I said in chapter 3.2., communicative action stops functioning as a *mode* of interaction. One must, it seems to me, keep apart these two levels of (1) individuals carrying out their individual goals and (2) doing so with (or without) regards to others' goals. One might speak of intentions concerning (1), and of an orientation with regard to (2) I think, as it refers to either communicative or strategic action. While one can surely act with the intention to reach consensus, it does not seem like one has to have such an intention to act communicatively in order to so. Consequently, for Dallmayr's argument to be successful, one would have to be arguing that while in pursuit of an individual goal, one has the intention to take others' interests into consideration while so doing.

---

[131] Seyla Benhabib, *Situating the Self*, p. 30.

This way of putting it is a bit odd however, as it makes the 'taking someone's interests into consideration (or disregarding them)' seem like the primary activity at some point in the interaction sequence, rather than the pursuit of one's own interests. If one were to do so, then one would be back to speaking of Habermas as oscillating between communicative action sometimes being a mode of interaction and sometimes reduced to actors' intentions (see chapter 3.2.). As I have already said, this does not seem like a good interpretation of Habermas.

What is interesting to me, is that if Benhabib is correct, her argument, that there are several plausible interpretations of the post-conventional stage of moral reasoning, has similar consequences as Hare's argument against the possibility of giving a complete justification of a moral norm, judgement, or principle.[132] This similarity is increased by Benhabib rejecting Habermas's idea that discourse serve to form agreement.[133] One must simply start from somewhere. Different persons adopt different premises due to living in different contexts, while Habermas would not disagree, Benhabib seems to take it further in the sense that this affects even one's interpretation of the content of the post-conventional stage – it affects one's understanding of impartiality, reversibility, and so on. My point being that if Benhabib is correct, much of the ground on which Habermas's theory is built will fall. As the theory of communicative action reflects his description of the post-conventional stage of moral reasoning, and the necessity of communicative action relies on it being the only reasonable one, if this is cast as a matter of acculturation, and there is no need for consensus either, then neither can the necessity nor the priority of communicative action be presupposed. If this is the case, then reason cannot be that closely connected to communicative action either, and it explains why one cannot be rationally bound to act communicatively.

Indeed, this is not much of a vindication of Dallmayr's argument; there still can be no oscillation between communicative and goal-oriented action. However, if one assumes that Benhabib is correct, and that Hare's conception of reason is that which Habermas calls cognitive-instrumental, and that there are no reasons to assume that one cannot rationally justify a normative statement using such an understanding of reason, then the distinction between communicative and strategic action does seem superfluous. As could be seen in chapter 3.1., even without this distinction, manipulation and deceit and force is rendered irrational and morally impermissible. Seen from this perspective, the two modes of action do seem to conflate, and does so without much damage to the rational justifiability of moral propositions.

---

[132] As Jenny Ehnberg suggests in her doctoral thesis, Benhabib's cognitivism is not very cognitivist since as to her, moral statements are not truth-apt. See Jenny Ehnberg, Globalization, Justice, and Communication, p. 161.

[133] Seyla Benhabib, *Situating the Self*, p. 37; Jenny Ehnberg, Globalization, Justice, and Communicaiton, p. 164.

5.3. Conclusion

The conclusions that follows from this study, it seems to me, is first and foremost, that Habermas is wrong in his claim that rational justification of normative propositions is impossible for a non-cognitivist. There are three main reasons for this:

(1) Habermas claims that action, when taking the perspective of an observer, becomes strategic since the only reasons available, are empirical one's. If reason is restricted to the cognitive-instrumental dimension, the only criteria for considering the rationality of an action is its effectiveness; it is a rational action if it is likely to succeed. Hare gives a good reason to doubt this argument as his version of the universality principle has the consequence of rendering some actions, which certainly might have been successful, morally impermissible.

(2) Following Habermas's linguistic philosophy, the rationality of an imperative request relies, similarly to a goal-oriented action, in whether one can expect the hearer to yield to the speaker's will. To be able to rationally justify a moral proposition, one has to be able to rationally motivate a hearer, not threaten them with sanctions or promise them rewards or manipulate them using lies or similarly. According to Hare, moral judgements is a form of imperatives, and as such, if one should believe Habermas, they cannot be rationally justified. As has been shown, however, this is too simplistic. An actor may comply with the command – or disobey it – as long as their doing so does not violate the version of the universality principle Hare proposes; if the actor can accept the consequences of their action were the roles to be reversed. It is true however, as Habermas claims, that while Hare's theory takes argumentation into consideration, it is not necessary to engage in it.

(3) Benhabib's argument is important to consider here too. Assuming that she is correct in that Habermas cannot assume, as he does, that there is but one correct description of the post-conventional stage of moral reasoning, then neither can he claim that any rational person must accept the universality principle. The consequence of this argument is that the possibility of the social world to come together as one becomes severely limited, since, as with Hare's argument, the only correctness conditions there is are relative to the premises from which one is reasoning, premises that are decided upon arbitrarily due to one's subjective perspective.

The necessity of cognitivism for the discourse theory of morality is another issue. Here, the evidence points in quite different directions. There are two dimensions to this answer: whether the discourse theory of morality is necessarily cognitivist in the sense that its key claims could not be sustained without cognitivism, or whether cognitivism is needed for this theory to rationally justify a moral judgement.

(4) The focus on rational agreement is probably the strongest aspect pointing towards the necessity of cognitivism for Habermas's theory. Having rational agreement as a goal to which discourses strives, it seems reasonable to call the theory cognitivist due to the focus on the "coming together" of the social world in a similar sense as the objective world is coming together.

(5) The social world would be rather superfluous in a non-cognitivist version of the theory more similar to Grenholm's category called moral experience. The important difference being that the social world seems like something which participants should reach agreement about, which is not necessary with regards to moral experience. Experience is something subjective; although actors might certainly have similar experiences, differences in experience is not something one can argue about, except for any empirical aspects pertinent to a particular case (did the speaker actually experience which they claim to have experienced? Do they behave as if they feel what they claim to feel?). That which is experienced is certainly external, but experience is not, and hence, the consensus-oriented approach would not be a possible one. One cannot make one's own experiences someone else's. Thus, as was said in the previous paragraph, unless cognitivism holds, an important aspect of the theory must be reconsidered.

(6) On the other hand, these two paragraphs do not affect the possibility to rationally justify moral statements for the reasons mentioned in (1), (2), and (3). What this implies is that if "necessary" means that without cognitivism the entire theory falls apart, cognitivism is not necessary for the discourse theory of morality. If, however, "necessary" means that if this is changed, then some central aspect of the theory must be revised – in this case, that argumentation cannot be required and that there can never be any right answer to moral issues, then yes, cognitivism is necessary to the discourse theory of morality.

6. References

Alexy, Robert, *A Theory of Legal Argumentation* (Oxford: Oxford University Press, 1989)

Benhabib, Seyla, *Situating the Self: Gender, Community, and Postmodernism in Contemporary Ethics* (Cambridge: Polity Press, 1992)

Davidson, Donald, 'Belief and the Basis of Meaning', in *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), pp. 141–54

———, 'True to the Facts', in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 37–59

———, 'Truth and Meaning', in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 17–36

Edgar, Andrew, *The Philosophy of Habermas* (Chesham: Acumen, 2005)

Ehnberg, Jenny, *Globalization, Justice, and Communication* (Uppsala: Acta Universitatis Upsaliensis, 2015)

Ekenberg, Tomas, Lars-Göran Johansson, George Masterton, and Pauliina Remes, 'Berättigande', in *Kunskapsteori: En Introduktion till Vetande, Berättigande Och Sanning* (Lund: Studentliteratur, 2014), pp. 33–51

———, 'Språk, Värld, Sanning', in *Kunskapsteori: En Introduktion till Vetande, Berättigande Och Sanning* (Lund: Studentliteratur, 2014), pp. 91–124

Finlayson, James Gordon, *Habermas: A Very Short Introduction* (Oxford: Oxford University Press, 2005)

———, 'Habermas's Moral Cognitivism and the Frege-Geach Challenge', *European Journal of Philosophy*, 13 (2005), 319–44

Geach, Peter, 'Assertion', *The Philosophical Review*, 77 (1965), 449–65

Grenholm, Carl-Henric, *Etisk Teori* (Lund: Studentliteratur, 2014)

Habermas, Jürgen, 'Discoure Ethics: Notes on a Program of Philosophical Justification', in *Moral Consciousness and Communicative Action* (Cambridge: Polity Press, 1990), pp. 43–115

———, *The Theory of Communicative Action. Vol. 1, Reason and the Rationalization of Society* (Boston: Beacon Press, 1984)

Hare, Richard, *Freedom and Reason* (London: Clarendon Press, 1965)

Johnston, James Scott, 'Is There a Need for Transcendental Arguments in Discourse Ethics?', *Educational Theory*, 66 (2016), 719–34

Kant, Immanuel, J. B. Schneewind, Marcia Baron, Shelly Kagan, and Allen W. Wood,

Johan Elfström

    Groundwork for the Metaphysics of Morals (New Hampton: Yale University Press, 2002)

Lafont, Cristina, *The Linguistic Turn in Hermeneutic Philosophy* (Cambridge: MIT Press, 1999)

Levine, Steven, 'Truth and Moral Validity: On Habermas' Domesticated Pragmatism', *Constellations*, 18 (2011), 244–59

Lillehammer, Hallvard, 'Moral Cognitivism', *Philosophical Papers*, 31 (2002), 1–25

Power, Michael, 'Habermas and Transcendental Arguments', *Philosophy of the Social Sciences*, 23 (1993), 26–49

van Roojen, Mark, 'Moral Cognitivism vs. Non-Cognitivism', in *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2016) <https://plato.stanford.edu/archives/win2016/entries/moral-cognitivism/>

Schroeder, Timothy, 'Propositional Attitudes', *Philosophy Compass*, 1 (2006), 35–73

Stout, Jeffrey, *Democracy and Tradition* (Princeton: Princeton University Press, 2004)

Strawson, Peter F., 'Freedom and Resentment', in *Freedom and Resentment and Other Essays* (London: Methuen & Co. Ltd, 1974), pp. 1–25

White, Stephen, *The Recent Work of Jürgen Habermas: Reason, Justice and Modernity* (Cambridge: Cambridge University Press, 1988)

Wright, Crispin, 'Inflating Deflationism', in *Truth and Objectivity* (Cambridge: Harvard University Press, 1994), pp. 1–33

Young, Iris Marion, *Justice and the Politics of Difference* (Princeton: Princeton University Press, 1991)