Moral Norms: Conventions or Norms with Universal Authority?

Academy of Science, Oslo / Christel Fricke, CSMN / IFIKK / Universitetet i Oslo

Abstract

In this paper, I sketch an interactive process of conflict solution and moral learning which brings about agreements on social and moral norms as well as the motivation to act in accordance with them. The norms emerging from this process are justified in virtue of certain features of the process itself. This sketch is strongly inspired by what Adam Smith, in his *Theory of Moral Sentiments* (1759), says about human morality. Furthermore, the process as I reconstruct it bears interesting similarities to the process of convention building as it has been described by David Lewis. Against this background, I explore how social norms can be conventional and still have universal authority.

Introduction

In recent years, scholars have taken an increasing interest in Adam Smith's moral theory. Among the numerous topics of their debates is the question whether Smith is defending a relativistic view of morality or not: Does he, in his *Theory of Moral Sentiments*, simply describe the behaviour of those of his contemporaries who are considered as moral models and thereby develop a sociological and psychological analysis of habitual practices of the people of his own time, country, culture and class? Or does he develop a normative theory defending the claim that moral judgments are based on norms with universal authority, an authority independent of actual social practices? Relativists point out that Smith's famous impartial spectator, the moral judge, cannot draw on any standards for proper ways of feeling, willing or acting which are entirely independent of the habits and practices of the cultural group to which he himself belongs.[1] Their opponents draw on Smith's belief in all human beings being equal and therefore equally bound by the same moral obligations.[2]

Both readings can rely on textual evidence. Should we therefore conclude that Smith's *Theory of Moral Sentiments* is internally inconsistent? Such a judgment would be premature. Rather than accusing Smith of an inconsistency and put his work back into the dusty drawers of the history of moral thought, we should follow a suggestion made by Samuel Fleischacker and see the tensions between Smith's relativism and his universalism as a challenge for future research on moral matters.[3] In this paper, I shall try to meet this challenge by developing a thought experiment the object of which is an interactive process of moral learning. The conception of this process relies extensively on elements of Smith's moral thought. Against the background of this thought experiment, both the relativistic and the universalistic elements of the *Theory of Moral Sentiments* can be seen as parts of one overall coherent and internally consistent moral theory. However, this theory denies the truth of two views, inspired by Kant's moral theory, which are common in philosophical thought about moral matters.

The Moral and the Conventional

1 See for example Allan Gibbard 1990, p. 280, Fleischacker 1999, p. 41ff., Forman-Barzilai 2006.
2 See for example Stephen Darwall 2006, p. 43ff. and Otteson 2002.
3 See Samuel Fleischacker 2005, p. 125/6.

Social life is life in accordance with social norms and rules.[4] It is common to distinguish between two kinds of social rules, namely between those which are moral and those which are not. The latter are often called "conventional rules". Thus, Shaun Nichols describes the experimental findings of developmental psychology by using the moral-conventional distinction:

> In the psychological literature, the capacity for moral judgment has perhaps been most directly and extensively approached empirically by exploring the basic capacity to distinguish moral violations from conventional violations. … From a young age, children distinguish canonical moral violations from canonical conventional violations on a number of dimensions. For instance, children tend to think that moral transgressions are generally less permissible and more serious than conventional transgressions.[5]

Among the examples for moral rules young children recognize as such is the rule forbidding hitting and pulling someone's hair, among the examples for conventional rules is the rule forbidding talking out of turn or during story time. The distinction between these two kinds of rules has been confirmed in many experiments and is now considered as "robust". In the light of this moral-conventional distinction someone who claims that the so-called moral rules are just as conventional as any other social rules will be considered as a moral sceptic. However, the use of the notion of a convention as opposed to that of a moral rule or norm depends on two assumptions inspired by Kant's moral thought: on a pre-theoretical understanding of the conventional as arbitrary and dependent on local authorities on the one hand and on the other on the assumption that the authority characteristic of moral norms cannot arise from anything as contingent as facts about humans and the needs and desires they happen to have.

David Lewis has developed a theory of conventions that does not confirm the first of these assumptions.[6] And one can argue for a denial of the second assumption: The difference between the authority of so-called conventional and that of so-called moral norms is a difference in degree, not in kind. If we give up the still widespread belief in the truth of these two assumptions, the question arises how the social norms that are conventional in Lewis' sense of the term relate to moral norms. Can there be moral norms which are both conventional and have universal authority at the same time? I shall try and argue for an affirmative answer to this question. The main part of my argument will take the shape of a thought experiment which relies extensively on Adam Smith's moral insights. It describes an interactive process of moral learning in the course of which people agree on social and, in the long run, moral norms. These norms acquire their authority gradually in the course of the process which brings them about. This process bears important similarities to the procedures of successful coordination of actions by which conventions are established. The thought

---

4 Sociologists commonly understand social norms as rules backed by social sanction. The notions of 'social norm' and 'rule backed by social sanction' can be equally applied to many of the normative standards underlying a social practice. Their extensions are, however, not intrinsically the same. There may be social norms, namely moral norms with absolute authority, the existence of which does not depend on social practice or, in particular, social sanction. The authority of such norms would be independent of their being present in an actual social practice. Kant prominently argued for the existence of such norms. Sociologists – and an increasing number of philosophers – do not believe in the existence of social norms independent of a social practice. They try to reconstruct the common sensical distinction between moral and other social norms without attributing absolute authority to the latter. The conceptual distinction between 'social norms' and 'rules' is, accordingly, less important for them.

5 See Shaun Nichols (2005), p. 356.

6 See David Lewis (1969).

experiment will hopefully show how to accommodate universalistic and relativistic tendencies in Smith's *Theory of Moral Sentiments*.

The thought experiment

Four different stages of the interactive process of conflict solution and moral learning which the experiment explores have to be distinguished:

Stage (i): Claims about human nature
Stage (ii): The mediating role of the impartial spectator
Stage (iii): Institutionalizing social rules people have agreed upon
Stage (iv): The ideal of the fully impartial moral judge

In the following, I shall focus on the first two stages and will mention the last two only briefly.

Stage (i): Claims about human nature

In the beginning of this process, there are human beings who are equally provided with two basic emotional dispositions and motivational drives, namely selfishness and sympathy. Selfishness and sympathy are to be understood along the lines of Smith's anthropology.

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it. (TMS I.1.i.1, p. 9)

For a contemporary reader it is important to make explicit that the selfishness Smith is here attributing to humans is a natural disposition which in itself is free from anything objectionable. If humans had not been provided with this selfishness, they would not care for themselves and would be unlikely to survive. For this first natural human disposition Smith also uses the notion of "self-love".[7]

For the other disposition Smith is here attributing to human nature he will in the following mainly use the notion of "sympathy":

> Pity and compassion are words appropriated to signify our fellow-feeling with the sorrow of others. Sympathy, though its meaning was, perhaps, originally the same, may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatever. (TMS I.1.i.5, p. 10)

At first sight, Smith's "sympathy" seems to be some kind of empathy (the disposition of humans to become aware of and share other's feelings) or some kind of altruism (the disposition to care for others without expecting any advantage from it for oneself). One may further think of the idea that sympathizing with someone implies liking this person (for what she feels or does or is). But none of these interpretations catches the full sense of Smith's notion of sympathy, and the latter would even be entirely misleading. This notion, however,

[7] See for exp. TMS, I.i.2.1, p. 13 and II.ii.2.1, p. 83.

plays a crucial part for the future development of the thought experiment as well as for Smith's moral theory.

Both self-love and sympathy are basic emotional dispositions, dispositions to have certain feelings, feelings with a strong motivational drive. Our *self-love* makes it that we feel good or bad about ourselves and care for our survival and well-being. S*ympathy* includes a disposition to share other people's feelings, to care for them and to desire their sympathy. As beings provided with sympathy, humans are essentially social. They have a natural desire to live in a state of mutual sympathy or social harmony, to be moral and nice to each other. They can satisfy their natural social needs only if they live in societies the members of which have agreed to act in accordance with certain social norms. These norms are not provided by a transcendent being. Nor does their authority derive from a contractual agreement. They have to be constituted or found by the people themselves, and in order to agree on them, people have to go through an interactive process of social and moral learning. The driving forces of this process are nothing but the needs and desires humans naturally have.[8] It is a process of civilization, but because it is driven by these needs and desires, it is a natural process nevertheless. Nature and civilization are, for Smith, not in opposition against each other.

Furthermore, Smith's notion of sympathy includes a normative element. Humans do not blindly sympathize with each other. They make their sympathy with other people's feelings dependent on the *propriety* of these feelings.[9] Feelings can differ in kind and in degree. Which kind and which degree of feeling it is proper for a person to feel depends both on her present physical and emotional state as well as on the circumstances to which she emotionally responds. The proper emotional response to the loss of a beloved partner is deep grief; responding with a similar degree of grief to the loss of an umbrella would be improper. Thus, humans provided with sympathy will share the grief of someone who has lost a beloved partner. But they will refuse to share a similar grief when they respond to the loss of an umbrella. Given its dependence on considerations of propriety, sympathy is not only an emotional disposition and a faculty of subjective evaluation by standards of personal liking and disliking. It is also a faculty of evaluation by shared standards of propriety. And the objects of sympathetic evaluation are not only feelings but also volitions and actions. This is because feelings are emotional drives, and Smith understands the propriety of actions partly in terms of the propriety of the feelings that motivated them.[10]

What are the standards of propriety? How do they arise? What is the source of their authority? How do people get to know them? And why are people motivated to feel and judge and act in accordance with them? The thought experiment is supposed to provide answers to all these questions.

Humans naturally provided with self-love and sympathy are involved in an interactive learning process. They are not by nature provided with any knowledge of how to satisfy their natural needs and desires; they have no a priori knowledge of what is good for them. And,

---

8 One may wonder how realistic Smith's picture of the needs and desires humans naturally have actually is. There is growing evidence from the side of both the behavioural and brain sciences that humans are not naturally driven exclusively by selfish concerns. These findings support Smith's picture. But this picture of the nature of the human mind and its basic emotional drives may well be incomplete; there may be emotional drives that are incompatible with the idea that humans naturally have a disposition to be morally good persons or nice people.

9 See TMS, I.i.1.6. – 12., p. 11/2.

10 Smith, however, does not deny that evaluations of actions also have to take the effects of these actions into account. See above, p. … and TMS, II.i.1., p. 67 and II.ii.3, p. 85 – 91.

according to the assumptions of the thought experiment, there is no one who can teach them what is good and what is bad for them. They have to find out, and as there are no teachers they can only do so by relying on the method of trial and error. For example, when they feel either hungry or thirsty, they have to find something to eat and drink. As they have to rely on trial and error, some may well not survive their experiments; but those who do will start learning what they can consume in order to satisfy the needs arising from their self-love. Similar learning processes take place when people try to satisfy any other needs – selfish and social – besides hunger and thirst. In the beginning, people will not have any idea of what it means to hurt someone or to hinder someone from satisfying the needs originating in her self-love. But they will be told by those whom they hurt or hinder from getting what they want. They will be exposed to the resentment and explicit complaints of these others, and they will suffer from a lack of sympathy from their side.

Learning is learning from experience, and this experience involves in particular the experience of failure. Failure to provide for one's well-being leads to weakness, illness and death. Social failure causes resentment from the side of other persons with selfish interests; it originates in overly selfish behavior. People who are provided with sympathy have a natural disposition to mind the resentment of others and therefore try to avoid it. However, they do not try to avoid it at whatever costs for themselves. They do not exclusively focus on pleasing others at the expense of their own well-being. What they have to learn in order to satisfy both their selfish and their sympathetic needs and desires, is to find a balance between the two: to satisfy their selfish needs and desires and to preserve the sympathy of the others at the same time. Those who succeed in finding this balance and enjoy mutual sympathy with all others have successfully learned to feel and act in the proper way.[11]

Stage (ii): The mediating role of the impartial spectator

So far, the learning process has been described as a process of interaction of individuals with the world and of individuals with other individuals where the interacting individuals are in some cases conflicting parties. But this is not the full scenario of the thought experiment. In addition to the conflicting parties (typically an agent and his victim), there is the impartial spectator, and the appearance of this character marks the second stage of the thought experiment.[12] In the framework of this experiment, two notions of the impartial spectator have to be distinguished: There is on the one hand the spectator in the sense of a mediator between conflicting parties, helping to find a solution of their conflict.[13] And on the other hand there is the spectator in the sense of an ideal judge of the propriety and impropriety of feelings and actions. The ideal judge either knows what is proper to feel and do under circumstances of various kinds; then, he has an objective notion of what is good and bad, he has obtained moral knowledge and can make the standards of his moral judgments conceptually explicit. Or, in cases of moral judgments the standards of which have not been made explicit as rules, he just relies on his feelings of sympathy and trusts these feelings.

---

11 Smith himself does not stress the procedural and interactive character of moral learning. However, the idea of an interactive process of moral learning is implicit in his theory of conscience development, and he explicitly refers to the role of learning from "habit and experience". See, for exp., TMS, III.2.2. – 3., p. 135.

12 See TMS, I.i.4., p. 19 – 23.

13 Smith himself does not make this distinction explicit. However, attributing a mediating function to the spectator as Smith describes his social role is not far-fetched. See for exp. TMS, I.i.1.7., p. 11 and III.2., p. 129. I would like to thank Samuel Fleischacker for drawing my attention to the passage on p. 129.

In the early stages of the interactive process of moral leaning there cannot be an impartial spectator in the sense of the ideal judge. The role of the spectator is that of a mediator between conflicting parties. His impartiality is, in a sense, negative: It consists in nothing but his not being himself involved in the conflict; his own selfish interests are not at stake under the given circumstances. Whereas the conflicting parties emotionally perceive the circumstances in which the conflict arose under a veil of partiality arising from their overwhelming selfishness, the spectator sees these circumstances, including the conflicting parties in their actual states, in a more neutral way.[14] His impartiality allows him to put himself into the shoes of both conflicting parties respectively and to try and understand how they experience the conflict. He will then try to suggest a way to overcome the conflict. As long as the spectator is just a mediator between conflicting parties, the only standards he has at his disposal for judging the propriety or impropriety of the feelings and actions of the conflicting parties are those he has himself acquired in his own former experiences as either and agent or a victim. Thus, the standards of propriety used by a spectator with only negative impartiality are culture relative, they cannot rightly claim universal authority.

Where the conflicting parties agree on a solution of their conflict with the help of the mediation of an impartial spectator, the three parties involved reach a state of mutual sympathy. An agreement between a very limited number of people, typically two conflicting parties and a spectator or mediator, does not as such accord with standards of propriety which can rightly claim universal authority. The standards emerging from such an agreement between just a few people are social norms endorsed by just the people who have actually agreed on them. But this very limited factual authority does not exclude them from being justified – even though their justification is of a modest kind. They derive the legitimacy of their authority for the people who have agreed on them from the procedure by which these people actually found the solution of their conflict. In the course of this procedure the conflicting parties have successfully overcome the partiality of their respective views of the circumstances under which the conflict between them arose. The standards of their agreement are, if not objectively impartial, at least more impartial than the entirely partial and mutually incompatible standards on which the conflicting parties originally relied.

Partiality and impartiality of standards of propriety come in different degrees. Due to a natural disposition to emotionally respond to circumstances in the world in an overly selfish way, people's feelings, judgments and actions tend to be partial. But as people are provided with both self-love and sympathy, they also have a natural desire to overcome their partiality and feel, judge and act in accordance with impartial standards of propriety. Any agreement between people on social norms that has emerged from a process in the course of which originally conflicting parties have found a solution of their conflict with the help of a mediating spectator is *proper* in terms of standards which are not as partial as the standards arising from the selfishness of the respective parties alone. These standards or norms can make claims to at least a limited degree of impartiality, and their correspondingly limited justification is a result of the procedure of finding a conflict solution. None of the conflicting parties has exercised his power over the other and imposed a procedure according to his personal preferences. The mediating spectator has not blindly taken the side of any of the conflicting parties – though the spectator may have found one of the conflicting parties entirely in the wrong. Overcoming a conflict with the help of a mediating spectator does not

[14] Whereas the metaphor of the "veil of partiality" is mine, the idea that people by nature tend to be more selfish than they should, that they have to learn to moderate their selfish feelings and that they can only do so with the help of others who object to their overly selfish feelings and behaviour, is explicitly stated in TMS. See, for exp., TMS, I.i.5.5., p. 25; II.ii.2.1., p. 82/3; III. 1.3. – 7., p. 110 – 113.

mean that the spectator makes the conflicting parties meet half way between their opposite standpoints.

How can people move from local agreements on social norms to universal agreements on social norms with unlimited, universal impartiality or objective propriety, that is, to an agreement on social rules that can rightly claim to be moral? They just have to continue with the business of interacting and finding solutions for conflicts with the help of a mediating and increasingly impartial spectator. They have to bring in more and more people whose points of view are taken into consideration. The more people are involved in a process of mutual adjustment of interests, the more impartial the resulting agreement and the corresponding social norms will be.

In the beginning of this process of moral learning the people involved do not share any notions of what is a socially acceptable or proper or good action and of what is a socially unacceptable, an improper or bad action. They acquire notions of what is best that are subjective or partial, and care only about the sympathy of those few others with whom they have the opportunity to interact. But after having gone through processes of mutual adjustment of interests and agreement with the help of a mediating spectator, shared notions of socially accepted or good actions arise. People will learn to take the interests of more and more other fellow creatures into account and social norms arise about what to do under circumstances of a certain kind. These norms can be used for purposes of conflict solution. But they can also function as norms the respecting of which allows people to avoid conflicts. This process of learning how best to satisfy selfish and sympathetic interests is a process of continuously objectifying the originally subjective notion of what is best to be done. What is objectively best is at the same time what is proper from a strictly impartial point of view, the point of view all people can share.

Stage (iii): Institutionalizing social norms people have agreed upon

At some point in this interactive process, the people involved will transform some of the norms of socially acceptable or good behavior they have agreed upon into positive laws. Political institutions will be created which then have the power to impose the common norms on every individual member of the society and punish any violation of these norms. The rise of political institutions characterizes the third stage of the process of moral learning. However, even at this third stage there will be the need of an impartial spectator as a mediator. Thus, even in a society the members of which have reached this third stage of moral development and moral learning, the impartial spectator's mediating function will not become superfluous. One can therefore draw the conclusion that interactive processes as those characteristic of the first three different stages of the thought experiment will, in the framework of a real society, take place at the same time.

It should be mentioned that Smith himself, in order to develop his moral theory, does not rely on the method of a thought experiment. He relies on the description of morally relevant behavior and of procedures of moral judgment that he could observe among his contemporaries. And as he could find evidence for human behavior and interaction as characteristic of the three different stages of the thought experiment, it can hardly be surprising that he did not stress the processual character of moral development and moral learning. Smith also seems to hold the view that not only is there no need for positive laws for all particular cases of morally relevant decisions, but that an attempt to impose more laws

than strictly necessary for the well-functioning of a society would be an illegitimate restriction of the individual freedom of its members.[15]

Stage (iv): The ideal of the fully impartial moral judge

The interactive process of social and moral learning described in this thought experiment has an ideal end: At its ideal end, all people have agreed on norms of proper feeling, judgment and action for the most relevant types of the circumstances under which conflicts may arise. We have to understand the "all" in "all people" as inclusive as possible: It includes not only all people living on earth at a given period of time but also all members of all future periods of time. Unless there is an end of the world and the life of human beings in it, this end will never be reached. Thus, the impossibility of the final agreement is not only pragmatic in kind, it is intrinsic.

However, the social norms all people will have agreed on at the end of humanity on earth will have universal authority. This universal authority arises from the all-inclusive number of the people who have agreed on them as well as on the nature of the process of agreement: No one has imposed any norm on anyone else the respect of which would have been to the advantage of one party exclusively. All have participated in this process as equals, all have equally contributed to defining the impartial social norms, trying to save as much as possible of the various selfish interests they have. The world society the members of which respect these objectively impartial social norms enjoy social harmony and overall mutual sympathy – they are all as happy as a human being can ever be. They are all fully impartial moral judges both of all others and of themselves.

At the ideal end of the process of moral learning, all people will have become moral persons. They do not only know what the objectively proper, the moral norms are, they also want to feel and act in accordance with them. As they are humans, they still tend to have overly selfish emotional responses, but they will try and not let these partial responses have an impact on the way they judge and act. Furthermore, they have internalized the norms of moral judgment and therefore do not depend on the experience of the resentment of others in order to keep their overly selfish emotional responses under control.

Smith describes such a happy and morally developed society in the following terms:

> All the members of human society stand in need of each others assistance, and are likewise exposed to mutual injuries. Where the necessary assistance is reciprocally afforded from love, from gratitude, from friendship, and esteem, the society flourishes and is happy. All the different members of it are bound together by the agreeable bands of love and affection, and are, as it were, drawn to one common centre of mutual good offices. (TMS, II.ii.3.1., p. 85)

The ideal nature of the ultimate moral norms implies that real humans can never claim to have knowledge of them – their epistemic ambitions in moral matters have to be modest. But even though the ultimate end of social agreement cannot be reached, even though knowledge of the ideal moral norms is not possible for real humans, they should not be discouraged from trying to progress in their social interaction and move in the direction of this ideal end. Moral progress is possible, and as long as people interact and deal with disagreements which arise

[15] See TMS, II.ii.1.8., p. 81. I would like to thank Samuel Fleischacker for encouraging me to make this position of Smith's more explicit.

between them in the way described in this thought experiment, they can be sure of moving in the right direction. And the social rules they agree upon through procedures of conflict solution as described above will be morally justified, even though in a limited way. The limitation is due to the fact that an actual agreement between real people can never rely on the consent of all possible people; but no one can be blamed for this lack of moral perfection.

David Lewis' theory of *Conventions*

The process of interacting and social and moral learning as described in this thought experiment does not only draw on Smith's observations of moral behaviour, moral judgment and moral learning among his contemporaries. It also bears important similarities with processes of convention building as David Lewis has analysed them. Where people feel, act and interact under circumstances which do not already reduce the number of ways of action which all parties involved could approve of to one, where there is more than one way to determine what is the good, the proper way to feel and act under these circumstances, they are faced with a problem of coordination of actions. This is where an element of convention building comes into the process described in the thought experiment.

Conventions as defined by Lewis are regularities of behaviour of the members of a social group. These regularities are brought about by a successful solution of a non-trivial problem of coordination of actions. Problems of coordination arise where members of a group of people have a certain interest in common, an interest which they can satisfy only if they coordinate their actions. They have, so to say, a common aim, and in order to reach this aim they have to find a joint strategy. A very simple example for a coordination problem is the case of two or more people who wish to meet. Lewis uses this example himself.[16] In order to meet at a certain place at a certain time, people have to coordinate their actions. The main purpose of this coordination is to determine the time and place of the meeting. Where people coordinate their actions successfully, they reach an equilibrium.

For a given coordination problem there may be one or more equilibria, that is, there may be one or more strategies of how to successfully coordinate the actions of the people in question. If people want to meet, for example, there may be one or several points in time and space where they can get together. In the case of a coordination problem with only one equilibrium, the solution of the problem of coordination is trivial. However, where there is a coordination problem with a multiplicity of equilibria, the solution of the problem is not trivial; it depends on the people's succeeding to coordinate their actions and establish a convention.[17]

Each coordination-equilibrium or successful strategy to solve a coordination problem is – considered in itself – as good as any other. Thus, there is an element of arbitrariness involved in the establishment of a convention: people have to choose one out of several equally good strategies for coordinating their actions.[18] However, this arbitrariness of a convention should not be misunderstood. A conventional regularity of the behaviour of members of a group of people is not entirely arbitrary because it exhibits the solution of a problem of coordination. It does not depend on local authorities. Every convention fits the problem of which it is a solution. And as this problem arises from a certain constellation of circumstances, from a constellation of facts, its solution can only rely on a coordination of action in accordance to

16 See David Lewis (1969), p. 5 and others.
17 For his final definition of a convention see David Lewis (1969), p. 78.
18 See David Lewis (1969), p. 70

these facts. Each coordination-equilibrium represents one way to solve the corresponding coordination problem in accordance to the facts. The arbitrariness of a convention therefore relies exclusively on the choice of one equilibrium among several equally good alternatives.[19]

While conventions are arbitrary in only this one respect, their contingency is threefold. (a) Whether the members of a group do at all agree on a convention depends on their having interests in common for the satisfaction of which they have to coordinate their actions. That they have such interests is likely, but it is contingent. It is likely partly because the aim of a coordination of actions does not have to rely on some kind of cooperation, it can also rely on a procedure that allows the members of a group to keep out of each others' ways. (b) Furthermore, which particular interest(s) members of a group have in common so that they are motivated to coordinate their actions is contingent. (c) And finally, every convention is contingent in the sense that it is due to an arbitrary choice of one solution for a coordination problem among several equally good alternatives.[20]

Even though there may be a number of coordination-equilibria for a given coordination problem which are – considered in themselves – equally good, the arbitrariness of the choice of one of them over any other may be further reduced by the introduction of additional criteria of choice. These criteria can be pragmatic in kind, criteria of economy for example. In the case of people who wish to meet and for whom there is a very large number of coordination-equilibria, that is, possible spatio-temporal meeting points, the preference of some or one of these equilibria over any other may depend on such criteria. Some of the spatio-temporal meeting points may be easier or quicker to reach than others and some may provide better facilities etc.

A convention in the sense defined by Lewis is not in itself good or bad, it is not normative. There is no obligation for the members of a group of people to coordinate their actions and establish a convention. However, once a convention has been established among the members of a group of people, it gives rise to a social norm that obliges them to act in accordance with it. Any one who fails to act in accordance with an established convention thereby hinders not only himself but also every other member of the respective group to satisfy the interest that had originally led to a coordination problem and its conventional solution. Where conventions have been established among the members of a group, there is legitimate social pressure to act in accordance with them. Lewis speaks of conventions in terms of "socially enforced" norms: "one is expected to conform, and failure to conform tends to evoke unfavourable responses from others".[21]

How do people proceed in order to establish a convention? The straightforward answer to this question is: By making a corresponding agreement. However, Lewis points out that an explicit agreement is not the only possible way for the members of a group of people to establish a convention. Another way is by silent agreement brought about in a process of trial and error and through the slowly emerging mutual expectations and actions in accordance with them.[22]

---

19 Marmor already made this explicit, saying that the arbitrariness of a convention does not imply indifference. See Andrei Marmor (1996), p. 355.

20 See also Tyler Burge (1975), p. 254.

21 See David Lewis (1969), p. 97 – 100, here p. 99.

22 The establishing of a convention via trial and error and silent agreement on how to coordinate actions is Lewis' main focus of interest. This is because he wants to argue for the conventional character of any natural language. A natural language, however, cannot be the result of an explicit agreement between

Conventional elements in the interactive process of moral learning as described in the thought experiment

The learning process described in the thought experiment above bears important similarities to the process of silent coordination of actions via trial and error that leads to a conventional coordination-equilibrium. The experiment starts with the assumption that people have an interest in common: they are by nature interested in living in a state of mutual sympathy and social harmony. In order to satisfy this interest they have to coordinate their actions. The goal of this coordination can be described in terms of finding the proper way to feel and judge and act under circumstances of a particular kind. We can assume that the respective coordination problems do not allow for a trivial solution: The proper way to feel and judge and act under circumstances of a particular kind is, in most cases, not fully determined by these circumstances. Or, put in Lewis' terms: Given certain circumstances of action, there may be more than one coordination-equilibrium, more than one strategy to coordinate actions that meets the conditions of propriety and therefore can be accepted by all parties involved. And where there is no trivial solution of a coordination problem there is a need of a convention.

One can illustrate the conventional character of a coordination of actions that allows for social harmony with the help of the following example. A group of people shares a car. All members have contributed an equal sum to the acquisition of the car and they equally share the costs of keeping it. Now, the proper way to share the car seems to be to let them all have it at their disposal for an equal amount of time. But there are many ways to bring this about; there are many different coordination-equilibria. One after the other can have the car for an hour or for a day or for a week or for a month. But the people might prefer a less formal solution of their coordination problem, they might want to get together every Sunday afternoon in order to decide who will have the car at her disposal for which period in the course of the following week. Any solution they can all unanimously agree upon is a coordination-equilibrium, and the choice of one of them rather than another can be either entirely arbitrary (they might go for a lottery procedure) or reflect additional pragmatic criteria that arise from the different purposes for which these people need the car.

Many of the social habits and the corresponding rules and regulations that we commonly call conventional are not conventional in Lewis' sense of the term. They do not represent solutions of coordination problems. Some of these rules may have been imposed by local authorities, such as the above mentioned kindergarten rules forbidding talking out of turn or during story time. Other rules we commonly call conventional were originally invented by members of a limited social group who, by following these rules, intended to distinguish themselves from other people and underline the exclusiveness of their circle. These others then made an effort to follow these rules in order to try and make themselves members of the originally distinguished group. Many of the so-called rules of etiquette have originated in this way. Processes of coordination as Lewis understands them are, however, essentially socially inclusive, not exclusive.[23] Thus, Lewis' notion of a convention is egalitarian in spirit.[24]

---

people because, for making this agreement, they would already need a common language. Where there is no common language, there can be no explicit agreement about how to coordinate actions.

[23] See David Lewis (1969), p. 46. I agree with Marmor who said that rules of etiquette are not conventional in the sense defined by Lewis. But this is not an objection against Lewis' theory of conventions. Lewis himself explicitly said that he "did not undertake to analyze anyone's concept of convention". See Andrei Marmor (1996), p. 364 and Lewis (1969), p. 46.

Even though Lewis' definition of a convention does not coincide with the common sensical notion of a convention, there is the question whether social rules which are conventions in Lewis' sense of the term can ever rightly claim to be moral. Moral rules have characteristic features which conventional rules as defined by Lewis lack:

(1) Conventional rules are essentially contingent, moral rules are not.
(2) The authority of conventional rules depends entirely on there being a sufficiently large number of people in a group or society acting in accordance with them; as soon as this is not any more the case, the convention looses its authority and disappears. The authority of moral rules, however, does not depend on whether or not people act in accordance with them.
(3) Whereas the motivation of an agent to comply with a convention depends on his interest in the consequences he can expect to bring about, moral motivation is commonly understood as the motivation of an agent to do her moral duty for the sake of duty.

In the framework of the thought experiment developed above, these objections against the idea of understanding moral rules as conventional can be met in the following way:

Ad (1)
An ordinary convention is contingent in a threefold way: (a) whether or not the members of a population have a common interest they can satisfy only if they coordinate their actions is contingent; (b) if members have such an interest, it is still contingent what kind of interest it is; (c) conventions emerge from solutions of coordination problems that can be solved in various ways; the choice of one particular way over any other is arbitrary and thereby conventional. However, the social rules that people agree upon in the course of the learning process described in the thought experiment are not contingent in the same threefold way. The assumption that all people are by nature provided with sympathy implies that they have by nature one particular interest in common, the interest in social harmony. They can satisfy this interest only if they coordinate their actions. The solution of the respective coordination problem can be contingent only in the third sense. This contingency, however, is compatible with the claims of morality. Otherwise we would have to assume that, under all possible circumstances of action, there is one and only one way to act in a morally right way. This does not exclude that every set of social rules which rightly claims to be moral contains certain rules which are not conventional even in the third sense of the term. These rules may still be the products of successful coordination of action, if we assume that there was only one equilibrium for the respective problems of coordination. Candidates for such rules might be the rules forbidding killing, harming for no good reason, lying and promise breaking.

Ad (2)
The idea that there are moral norms and that their authority is universal even if there is no human being in the world who makes the slightest effort to act in accordance with them, even if there is no human being who even cares about them, presupposes an understanding of moral norms along the lines of the second of the above mentioned Kant-inspired assumptions. In the framework of the thought experiment developed earlier, this idea does not make sense.

---

[24] Burge has pointed to another discrepancy between the common sense of 'convention' and the sense defined by Lewis: Rules may be conventional even if the people following these rules are not aware of their conventional nature, they may not realize that there were other strategies for solving the underlying coordination problem. See Tyler Burge (1975).

Morality exists and matters only if there are people in the world who care about it. But the thought experiment entails that if there are mentally healthy people, they will, by necessity, care about morality, because they are provided with sympathy and care about each other. They will get involved in a process of coordinating their actions which, in the long run, will lead to an agreement on social norms which can rightly be considered as moral norms. The second assumption takes it that moral norms have universal authority, that the normative judgments based on them are true in all possible worlds. The universal authority of moral norms as justified along the lines of the thought experiment depends, however, on matters of fact, that is, on the existence of human beings provided with sympathy. There may well be worlds where no such people exist. But wherever such people do exist, they are bound by moral norms with universal authority. It will take them some time to learn how to spell out these norms in some detail and to establish the respective conventions, but this does not make the authority of the norms dependent on any contingent decisions about whether or not to care about morality.[25]

The process of moral learning as described in the thought experiment is not exclusively a process of finding out what the moral norms are. Otherwise these norms would be fully determined by the factual circumstances under which people happen to live. In the course of the learning process, people have to agree on the rules they want to follow, they have to make choices between different solutions of coordination problems. Therefore, this process is also a process of constituting social norms. The choices people have to make are choices between different and mutually incompatible sets of norms. But all these sets of norms are equally justified because they all represent solutions to coordination problems which originate in the natural interest in mutual sympathy and social harmony.[26] Once this process of moral learning has reached its ultimate ideal end, most of the morally relevant choices have been made and people will have agreed on one set of mutually compatible norms. The universal agreement on these norms depends at least in part on choices which are arbitrary in the sense explained above, and in this sense they can be called conventional. But this does not mean that they could still be changed. Therefore, these norms have universal authority and can rightly claim to be moral.[27]

Marmor seems to have a similar problem in mind when he points out that "one can be criticized for not complying with a moral rule, even if most others in his community fail to comply as well. One cannot be criticized, however, for failing to comply with a convention which is not actually practised."[28] A scenario of the kind Marmor describes here cannot occur in the framework of the thought experiment. Agreeing on a convention and complying with it is a natural need for all people provided with sympathy, the interests arising from this natural disposition are not contingent; therefore, the interest in following the conventions indispensable for satisfying the natural need for mutual sympathy and social harmony will always keep alive.

[25] Furthermore, the adherents of the second assumption will have to explain how people who are not even provided with sympathy can be motivated to care about morality and act in accordance with moral norms.

[26] The rules in the respective sets will be equally justified – otherwise the choice between them would not be arbitrary in the sense explained above. The difference between these sets and their mutual incompatibility does not exclude that there are certain rules which form part of every such set. Thus, it may well be that all such sets include a rule forbidding torturing babies.

[27] It seems to me that this is Smith's answer to the question raised by Andrei Marmor, namely how a rule can have normative authority and be arbitrary at the same time. See Andrei Marmor (1996).

[28] Andrei Marmor (1996), p. 357.

This, however, does not exclude that even in the framework of the thought experiment some individuals will sometimes refuse to comply with actual conventions. There may be two reasons for this: Either these people find themselves overwhelmed by their selfish needs and desires and are therefore driven to satisfy these without taking their social needs into account; in this case, the other people have a right to punish them for their unconventional behaviour. Or they consider the conventions people have actually agreed upon as imperfect and in need of improvement. They will then raise their doubts concerning the actual conventions with their fellow creatures. If they can give reasons for their doubts the others cannot deny, they will all engage in a process of improving the conventions on which they have actually agreed. Once the conventions have been changed in accordance with the doubts raised, all will be willing to comply with them. As long as the interactive process of convention building has not reached its ideal end, there will always be people who raise doubts about the actual conventions and try to argue for the need of further improvement. But they can raise these doubts only on the basis of their own selfish and social needs and desires. No one can raise such doubts on the basis of moral norms to which he or she claims to have privileged access – in the framework of the thought experiment, no one has such privileged access to knowledge of ultimate moral norms.

Ad (3)
The last of the above mentioned challenges a conventionalist understanding of moral norms has to face concerns its consequentialist implications for moral motivation. People agree on conventions and comply with them because they have certain interests which they cannot satisfy without coordinating their actions. Understanding moral norms as a special kind of conventional norms implies a consequentialist understanding of moral motivation: people are motivated to comply with moral norms because this is the only way to bring about certain effects, effects by which they can satisfy certain interests. It is, again, on the background of a moral theory along the lines of the second assumption mentioned above that this conception of moral motivation seems problematic. This assumption implies that people have to comply with moral norms whether they like it or not, whatever their interests happen to be, and whatever consequences are to be expected from the respective actions. What is excluded from moral deliberation and decision-making is a concern about the consequences to be expected for the individual well-being of the agent. The moral agent has to comply with moral norms for the sake of duty alone, not because she is interested in bringing about a certain state of affairs.[29]

There is then the question why adherents of the second assumption so vehemently defend a non-consequentialist understanding of moral motivation. First of all, they do not share the very optimistic view of human nature and of the basic motivational drives inherent in it on which the process described in the thought experiment depends. Furthermore, they think that moral motivation can by no means be made dependent on contingent facts. A consequentialist understanding of moral motivation, however, would make it dependent on a large number of contingent facts, including facts determining the success or failure of an action as well as facts concerning the behaviour of other agents. In a society most members of which do not habitually comply with moral norms, the moral agent is always running the risk of sacrificing his own well-being and letting others take advantage of him.

[29] However, not even an adherent of the second assumption would deny that moral actions would, if undertaken by a sufficiently large number of people, have consequences which would be advantageous for all people.

Adherents of the second assumption reject an understanding of moral motivation which is based on human needs rather than on the idea of an unconditional moral duty. But, in the framework of the thought experiment, the need from which moral motivation arises is a basic human need, present in all mentally healthy people. Peoples' well-being and happiness depend on the satisfaction of this need as much as on the satisfaction of the needs arising in their natural self-love. Trying to defend this conception of moral motivation, the question to be raised is this: What would, in the framework of the thought experiment, be possible sources of moral discouragement? In the real world, there are people who do not and probably never did care about doing what is morally right and who comply with moral norms only if it suits their selfish interests. There are also people who, in the course of their lives, give up caring about other people or sympathizing with them or caring about the sympathy of other people, people who become moral cynics. However, the phenomenon of moral cynicism is incompatible with the basic assumptions made in the framework of the thought experiment. Within this framework all people are equally provided with sympathy and share the ambition to coordinate their actions in such a way that they can all live in a state of mutual sympathy and social harmony. And the experiences they make in the process of coordinating their actions and agreeing on social conventions are never so discouraging that they will give up their natural moral ambitions.

Conclusion

Smith, in his *Theory of Moral Sentiments*, develops a normative theory by drawing on observation of the way people make moral judgments and interact accordingly. Whereas he is widely admired for his insights into human nature in general and into motivational psychology in particular, the question of how to understand his normative theory is more controversial. The thought experiment developed above is supposed to reconstruct a normative theory of moral judgment and the authority of moral norms that is in accordance with Smith's claims about human nature and the moral culture emerging from it. In the framework of this experiment, both the relativistic and the universalistic elements of *Theory of Moral Sentiments* can be accommodated:

Human nature is essentially moral. Mentally healthy humans have a natural interest in living in accordance with social norms which have universal authority. However, they are not by nature provided with a priori moral knowledge. For acquiring moral knowledge as well as the motivation to act in accordance with it, they have to go through an interactive process of conflict solution with the help of an impartial, mediating spectator. This process brings forth agreements on social norms the authority of which, in the early stages of the process, is limited to small groups of people. Such local agreements, I have argued, can be understood as conventions in Lewis' sense of the term. The respective social norms have limited authority because they are based on an agreement of a limited number of people. But this does not mean that they can in no way be morally justified. In so far as they represent successful solutions of coordination problems, they have justified authority for all those people who participated in finding this solution. The authority of social norms based on local agreements and the authority of social norms based on universal agreements differ in degree, not in kind. All social norms which are explicable in terms of convention building have justified authority in virtue of having been established by such a process. The degree of justification of any conventional social norms is a function of the number of people who have been involved in establishing them. Convention building is an essentially egalitarian process. The authority of conventional social norms does not depend on ad hoc authorities or the exercise of local

power. Nor should it be misunderstood in terms of the authority of a majority vote. It depends on the equal authority of all those involved in establishing them. Its reach is limited to this very group of people. In a sense, the authority of conventional social norms is internally general and externally limited – unless it is based on the authority of all people. In the latter case it is universal.

In the framework of the thought experiment, there is room not only for moral progress of a society over time but also for moral development in different societies which, independent from each other, are involved in establishing conventions and improving on them. Different societies may have made different choices of solutions of non-trivial coordination problems. And the more different the factual environments in which a group of people live are, the more the coordination problems they have to solve will differ. Thus, there is room for different social practices in different societies which may all be equally justified. Cultural pluralism and relativism of social practices with justified, even though limited authority, is compatible with an ultimately universalistic understanding of morality. The point is that moral authority and moral justification come in degrees.[30]

Bibliography

Burge, Taylor (1975), On Knowledge and Convention. In: *The Philosophical Review* 84/2, pp. 249 – 255.
Darwall, Stephen (2006), *The Second-person Standpoint. Morality, Respect, and Accountability.* (Cambridge/Mass and London/England: Harvard University Press)
Fleischacker, Samuel (2005), Smith und der Kulturrelativismus. In: Christel Fricke und Hans-Peter Schütt (eds.), *Adam Smith als Moralphilosoph.* (Berlin: deGruyter), pp. 100 – 127.
Fricke, Christel (2009), Passt der Mensch in die Welt? In: Heiner Klemme (ed.), *Kant und die Zukunft der europäischen Aufklärung.* (Berlin: deGruyter), pp. 292 – 318.
Forman-Barzilai (2006), Smith on 'connexion', culture and judgment. In: Montes, Leonidas and Schliesser, Eric (eds.), *New Voices on Adam Smith.* (London/New York: Routledge), pp. 89 – 112)
Gibbard, Allan (1990), *Wise Choices, Apt Feelings. A Theory of Normative Judgment.* (Oxford: Clarendon Press)
Lewis, David (1969), *Conventions*. (Oxford: Blackwell)
Nichols, Shaun (2005), Innateness and Moral Psychology. In: Peter Carruthers, Stephen Laurence, Stephen Stich (eds.), *The Innate Mind. Structure and Contents.* (New York, OUP), pp. **
Marmor, Andrei (1996), "On Convention". In: *Synthese* 107, p. 349 – 371.
Otteson, James R. (2002), *Adam Smith's Marketplace of Life.* (Cambridge: Cambridge University Press)
Smith, Adam (1759/1984), *The Theory of Moral Sentiments* (quoted as 'TMS'). (Indianapolis: Liberty Fund)

[30] I would like to thank the members of the Norwegian Academy of Science and the members of the Moral Philosophy Club at CSMN for very helpful comments to earlier versions of this paper. Many thanks in particular to Kari Refsdal, at present a PhD student at CSMN, who has helped me with the edition of the text.