Chapter 4

Meaning and measurement of moral competence

In this chapter I will describe an innovative, new measurement methodology, Experimental Questionnaires (Lind 1978; 1982), which allows us to measure *internal*, *structural* dispositions in an *objective* way. There are many methods of measurement and observation which have one or two of these features but not all three of them. Most, if not all, widely used psychometric tests of attitude and ability measurement are certainly objective; but they do not allow the measurement of internal, structural properties of a person's behavior. In contrast, there are many so-called *qualitative* methods, like clinical interview methods, which attempt to measure internal, structural dispositions but are not objective and, therefore, are received with skepticism. Their application and their scoring rest too much on the researcher's intuitive judgments. These judgments are easily biased by the theoretical assumptions which they should test. The *Moral Competence Test* (MCT), which I have designed as an Experimental Questionnaire measures people's internal, structural moral competence in a completely objective way. The MCT makes it possible to study moral competence in a scientific way.

"For a large part of the early 20th century it was generally held that moral phenomena cannot be studied scientifically," wrote Bill Kurtines and Jacob Gewirtz in their handbook article on moral development (Kurtines & Gewirtz 1995a, p. 3). The main obstacle, it appeared, was the lack of objective methods to measure moral dispositions, that is, moral orientations and moral competence. Many answer this question in the negative. They believe that internal morality cannot be measured at all. They declare moral dispositions to be unobservable, "latent" constructs that can only be assumed and inferred by means of intuitive methods, which are hardly explicable. Moral dispositions thus became a matter of belief rather than of scientific research. In contrast, I will show that they can be objectively measured using an individual's manifest behavior and that questions regarding the nature, development and education can be answered through scientific research.

A very famous attempt to measure moral dispositions is the method of clinical interview by Piaget and Kohlberg. The idea behind this method of measurement is to study the moral judgment of children (Piaget) and adults (Kohlberg) about the decisions of protagonists in hypothetical dilemma stories: it "is the moral judgment that we propose to investigate, not moral behavior or sentiment." (Piaget 1965, p. 7) From the body of interviewees' answers the researcher distills scores, which they try to fit into a well-developed theoretical scheme. This method gave birth to the notions of developmental schemas, phases and stages.

The clinical interview method has been criticized for several reasons, which I have summarized elsewhere (Lind 1989). Some cognitive psychologists even thought that "The interview method is doubly dangerous. It not only failed to see what was revealed by the integration method, but claims to see something quite different. [...] This failure of the interview method reflects the treacherousness of verbal reports for cognitive analysis." (Anderson 1991, p. 167)

Piaget and Kohlberg were quite aware of the ambiguities and limitations of their method. Actually they wanted to contribute to the understanding of human moral behavior. But how does people's verbal reasoning about hypothetical stories relate to behavior? Piaget admitted: "A great danger, especially in matters of morality, is that of making the child say whatever one wants him to say." (Piaget 1965, p. 8) "The point, then, that we have to settle is whether the things that children say to us constitute, as compared to the real conduct, a conscious

realization [...]. We do not claim to have solved the problem completely. Only direct observation can settle it." (p. 115)

Kohlberg also had his doubts about assessing moral behavior through interviews: "From the point of view of structural theory, a subject need not be self-consciously aware of his own stage structure." Nevertheless he insisted that "Structural consistency […] should be found only by a psychologist abstracting structural features from spontaneous responses." (Kohlberg 1979; p. xiv) He was caught in a methodological dilemma. On the one hand, he believed that in spite of their shortcomings personal interviews were more apt to tap a person's internal moral competence than "behavioristic conceptions of moral conduct," because these "typically define conduct as moral if it conforms to a socially or culturally accepted norm." (Kohlberg 1984, p. 392) On the other hand, Kohlberg envisioned quite another method of assessing moral competence or structure: "[Structure] is a construct rather than an inference, and is warranted only on the grounds of 'intelligible' ordering of the manifest items. One might say that the hypothetical structure is the principle of organization of the responses." (p. 408) Therefore Kohlberg asserts that "the test constructor must postulate structure from the start, as opposed to inductively finding structure in content after the test is made. [...] If a test is to yield stage structure, a concept of that structure must be built into the initial act of observation, test construction, and scoring." (Kohlberg 1984, pp. 401-402)

When I told him about our approach of assessing moral behavior and sentiment through direct observation, that is, through an objective test which did not require the participants to conform to social norms and be conscious of their "structural consistency", but would make the principle of organization intelligible through the manifest ordering of an individual's responses, Kohlberg joked that he had such a test already in his drawer. Indeed, Kohlberg laid out very clearly his alternative (never realized) vision of structural measurement of moral structure already in his *Heinz-Werner-Memorial Lectures* (Kohlberg 1981). He acknowledged the MCT's capacity to "assign a pure structure score for an individual […] I believe," he wrote, "this to be a highly promising approach." (Kohlberg 2010, p. xvi, originally published in 1985)

We developed our approach before the publication of his lectures, drawing from different sources of experimental psychology (Lind 1978; 2008). Yet, some earlier published ideas were decisive for the development of our Moral Competence Test (MCT), which I describe in this chapter: Kohlberg's notion of moral (judgment) competence attracted me to the field of moral psychology and education and also triggered my search for an adequate objective methodology. His rejection of "defining individual morality as behavioral conformity to the more common rules of the individual's culture," (Kohlberg 1958, p. 2) has motivated me to search for an alternative to mainstream psychological test theory and to design a new methodology for measuring *internal, structural* dispositions, called *Experimental Questionnaires* (Lind 1982). With Experimental Questionnaires we have anticipated his requirement that the measured concept, moral competence, should be built into the initial act of test construction and scoring.

4.1   What is our concept of moral competence?

Earlier, in simple societies, it seems, we needed no special skills to be moral. It was enough if we accepted and obeyed all the rules on what and what not to do. The difficulty was 'only' to notice and remember these rules as they were so extensive. But even here one was challenged when the rules came into conflict. Even from early times in human existence one was not allowed to kill others. But if someone threatened to kill another person then this rule had to be

overridden. People dissolved such problems by allowing the validity of such a rule to be limited to their own family, their own tribe, and later to their own nation. Whoever did not keep to these rules was ejected and thus lost the protection of these rules. They had "forfeited their rights." The application of these rules often led to violent confrontations and wars.

During the Enlightenment era, the idea of universally valid rules or principles picked-up steam in philosophy and also in everyday human thought. Kant, one of the most important representatives of the Enlightenment (Neiman 2009; Höffe 2012), defined morality no longer merely as a list of prohibitions and commandments for behavior, but rather as compliance with universally valid moral principles. By this he meant those maxims of action which we would wish to be universally valid and applied. According to this categorical imperative, as he called it, our behavior is thus only moral if it coincides with our principles. At first glance, this new ethic of the Enlightenment had a great advantage from a psychological and educational point of view. We need to remember but a few rules.

Philosophers like Kant believed that the observance of these rules required at most motivation and courage ("Sapere aude!" – Dare to think!), but no particular moral competence. But is that true? Consider the thought experiment which he made to demonstrate the universal validity of the moral principle of truth. He argued that the principle of truthfulness must always apply, even if one endangers a person's life. Thought experiments like this are useful to clarify a philosophical point but they are certainly not apt to help us in every-day decision-making. In everyday life, we must face the fact that precisely because principles claim universal validity we are repeatedly brought into situations where we have no other choice than to violate one or more of our precepts.

Many philosophers appear not to have seen this problem. An exception that stands out for me is the approach of communicative ethics formulated by Jürgen Habermas and Karl-Otto Apel (Habermas 1990; Apel 1990). As far as I can see, they have not directly grappled with the problem of moral dilemmas, but their approach provides a good basis for tackling dilemma situations. Their ethics claims, roughly speaking, that we should solve moral problems and conflicts by reasoned weighing of arguments, that is, seeking for solutions through thinking and discussion, and not through violence – against ourselves or others – and not by self-deception, fraud, nor by the exercise of power. Habermas defines moral reasoning as "the eradication of those power relationships that are embedded inconspicuously into communication structures, and that prevent the conscious conflict resolution and consensual conflict regulation through intra-psychic as well as human blocks of communication." (Habermas 1976a, p. 34) Moral competence is therefore needed to overcome moral dogmatism or rigorism – that is, the one-sided following of one moral principle at the expense of another. Dogmatism is also a form of violence and often leads to overt violence or even terrorism. This, it seems to me, is a moral imperative that is unconditionally universalizable and fully coincides with both the external standard and internal morality: develop everyone's moral competence!

Moral competence is still a relatively new area of psychological research and applied education. Long ago, Socrates spoke of the "ability to do the Good" and Darwin (1966/1871) used the concept of moral ability. Max Levy Suhl (1912) found that criminal youths lacked not moral principles but the ability to apply them (which he called *moral maturity*). But only with Piaget's studies on the moral judgment of the child has this thread on "moral maturity" been systematically adopted. Kohlberg, whose work, built on Piaget, defined moral judgment *competence* as "the ability, to make decisions and judgments which are moral (i.e., based on internal principles) and to act in accordance with those judgments." (Kohlberg 1964, p. 425)

Comparing this definition with the common definition of morality at the time (and even today) as conformity of behavior with external rules and norms, it is quite remarkable. It contains three far-reaching changes:

- that morality has an ability aspect and can therefore not be simply reduced to its affective side,

- that morality is defined as conformity to *internal* principles, and only then can give meaning to concepts like moral ideals, autonomy and responsibility (Pittel & Mendelsohn 1966),

- and that action needs to be part of the definition of moral competence. That is, we are talking only of a competence when it manifests itself in behavior, and does not only exist as an idea or intention.

But Kohlberg's definition lacks reference to the difficult moral task of coping with dilemmas. For Kohlberg, solving moral conflicts and problems does not seem to require any special skills, but only a decision of the will, which involves feeling responsible for the implementation of a moral principle (Kohlberg 1984, pp. 498-582). The early work on his measurement instrument, the *Moral Judgment Interview* (MJI) makes positive efforts in this regard, but these efforts have fallen prey to later revisions, probably because their importance had not been recognized by the authoring collective (Colby et al. 1987).

Thus it should be clear that in real life we cannot confine ourselves to the question of whether people should act in accordance with their moral principles, but we also have to address the question whether they are also capable of dealing with the inevitable conflicts between these principles, that is, whether they are able to weigh conflicting thoughts and communication with like-minded people as well as with opponents. If we were to limit ourselves to measuring and teaching the enforcement of individual precepts, we would not promote moral discourse and peaceful conflict resolution but morally motivated violence. Eventually research has shown that the higher people's moral competence is, the more clearly they accept shared moral principles and reject low-type moral orientations (see page 74).

4.2 How to measure moral competence

The concept of moral competence which we have just described suggests that we measure it according to its definition, namely as the ability to engage in moral thinking and discourse. Today we can laugh at the problems we had to overcome in order to turn this concept into a valid measurement instrument. When you look at the product of our endeavors, you might exclaim: Oh, that simple! But as simple as this test looks on a first glance, its realization was not easy at all. When, in the 1970s, my colleagues and I sat down to create an objective test of moral competence we had to answer three important questions for which no answers were available at that time:

- First, how can we test a competence by the internal moral standards of the participant rather than by external standards? Testing abilities or competencies by external standards was (and still is) very common in psychological measurement, but it is completely at odds with the philosophical definition of morality that we have adopted (Pittel & Mendelsohn 1966; Kohlberg 1958; 1981; Lind 2008b).

- Second, what would constitute an adequate moral task?

- Third, how can we disentangle the ambiguities of isolated responses by participants and validly assess moral competence? We all know that an isolated act usually does not allow

us to make inferences on a person's underlying dispositions. When people greet us in a friendly way it does not necessarily mean that they like us; it could also mean that they are polite people, or that they want to hide their plans for an ambush on us. Similarly, when people agree with our principle moral argument about an issue, they may do so for various reasons: because they really appreciate moral principles, because they like to back up their (otherwise irrational positions) with moral language, or because they may just agree with everything a certain authority says. Classical approaches to measuring moral orientation and competence do not allow us to disentangle the multiple factors that usually determine an individual's response, but base their measurement on two doubtful assumptions: (a) that all responses to a moral test are determined substantially by the hypothesized disposition, and (b) that all variations of an individual's responses to test questions reflect pure random measurement error. The first assumption has been refuted already by Greene (1954) for attitude tests in general, and by Wuttke (2007) for scholastic ability tests. The second assumption was questioned by the results of my secondary analysis of longitudinal data on the political attitudes of college students (Lind 2010b). Indices for "measurement error", "reliability" and "internal consistency" showed strong trends which could be explained better by changes in the cognitive structure of the participants than by pure random error. Why would such indicators vary and change so much across studies if they really were only an attribute of the measurement process?

So what was actually needed was a new kind of measurement that is consistent with the assumptions of modern moral psychology (Chapter 3, page 51). Moreover we felt the need to create an objective test which was transparent for anyone who doubted our findings. In other words, we needed a completely new measurement theory to produce a theoretically valid test of moral competence. We therefore developed from scratch a new test, taking our bearings from the methods of experimental psychology (Lind 1978; 1985a; 2008b).

The incompatibility of classical psychometric methods with moral psychology

When we looked for suitable methods for measuring moral competence in the 1970s, we found only the clinical interview methods by Piaget and Kohlberg. Piaget's method was not really intended to measure an individual's moral dispositions but to assess general principles of moral development. Kohlberg's *Moral Judgment Interview* (MJI) came closest to our concept of moral competence. But we decided against it for two reasons. The one reason was practical. The MJI was not only too expensive for application and evaluation for use in our research project of the time, in which about 5000 students were to be interviewed on multiple occasions in five European countries over eight years (Bargel et al. 1982; Lind 1986a). The second, and more important, reason had to do with the measurement theory which forms the basis of the latest version of the MJI (Colby et al. 1987). Measurement theories are not a neutral statistical method, as many believe, that "merely" help "to reduce large amounts of data," and to make it presentable. Measuring theories contain hidden content-based assumptions about the nature of the measurement subject, which are not always compatible with our knowledge and assumptions about moral competence. Almost all tests in the social sciences are based on two implicit psychological *assumptions* (Gulliksen, 1950):

• First, we have the assumption that the ability to be measured is revealed in *every single* question or task in a clear and unambiguous manner. That is, it is assumed that the answers to each question of a questionnaire (item) – for example, the respondent's judgment about a particular argument – is caused completely by the disposition which is measured (in this case the moral competence) except for a "measurement error". Newer

measurement theories – such as the *Rasch Scaling* approach and the *Item Response Theory* – only refine this assumption, but do not call it into question.

- Second, statistical models for the design and evaluation of tests are based on the assumption that the measurement error is purely random and can be compensated for by a simple trick that one can copy from measurement of physical objects: in order to reduce the measurement error, simply measure several times and then calculate the mean score.

Both assumptions of classical test theory are wrong from a psychological point of view:

- The first assumption is in any case contrary to the definition of moral competence. The individual judgment or act is not always determined by only one factor, but usually by several simultaneously. As Kohlberg noticed, every judgment, each individual answer and single action can be, and usually is, determined by many factors simultaneously because "The solution [of a moral dilemma] must do justice both to what the self believes and yet meet the situation. Thus the choice is difficult in the sense [...] of doing justice to all the values which the self believes are true and important." (Kohlberg 1958, p. 128) How then should we be able to draw conclusions about a respondent's moral competence based on a single answer? A competence can be assessed only if one considers the situation, that is, the context of the task, and if one can rule out the possibility that the answer is determined by other conceivable causes (Wakenhut 1978). These factors cannot simply be regarded as measurement errors, as classical test theory assumes.

- Even if the deviations in answers had to do with purely random "measurement error" in the statistical sense, one could not so easily reduce them, as is generally possible in the measurement of inanimate objects. In humans each measurement leads to reactions which affect the next measurement. You cannot repeatedly ask people the same question as would be necessary to reduce measurement error. Therefore, new questions are used for repeat measurement in all tests in place of the same items. But there are hardly any questions that are different and yet measure the same thing. Therefore, the rule is: the longer the test, the lower its validity. That is, the longer the test, the more accurately it measures something, but unfortunately we do not know what it actually is.

- Conventional tests are not suitable to measure structural properties of human functioning. Structures underlying behavior show up only in the individual *patterns* of a person's behavior. But conventional tests focus on isolated responses to test-questions, thus ignoring the relational properties between an individual's responses. A simple example shows how misleading this can be: when we ask people how they feel, their answers are taken as a sign of a certain general disposition (like *happiness*); when we repeat this question, classical test theory interprets this as an identical measurement to check on the "reliability" of the question. But everyone knows that changes in the response can have a number of good reasons which the test simply ignores, namely a change of happiness, or a change of interviewer (a stranger often gets a different answers than a friend), or another good reason. Classical measurement approaches simply lump those structural properties under the assumption of "measurement error" und thus preclude deeper psychological analysis.

In the same way the basic assumption that all human dispositions are only "hypothetical" is opposed to the idea that a competence cannot be considered real unless it is demonstrated in overt behavior. Competences can "only be determined for certain based upon their tangible manifestations, i.e., based on performance phenomena." (Habermas 1990, my translation of the German edition, 1983, p. 198) That is, if moral competence means that an individual's behavior is truly oriented to moral principles, then it is not enough if we only capture their

moral orientation (attitudes, values, etc.). We must also analyze the relation between attitudes and behavior.

Unfortunately the inadequate test theory continues to dominate moral psychological research.[7] Its proponents reacted vehemently against Kohlberg's structural approach to the measurement of moral development (Kurtines & Greif 1974). At the outset Kohlberg had distanced himself from classical test theory. In his dissertation he chose instead Piaget's clinical interview method because he thought it would allow him to assess people's internal moral structures better (Kohlberg 1958). Later, he gave in to the criticism from orthodox psychometric theorists. He and his associates increasingly changed the Moral Judgment Interview to align it with the requirements of classical test theory (Colby et al. 1987). Thereby, it gained reliability but, I believe, lost validity. I fully agree with John Broughton, who countered Kurtines and Greif's critique: "Structural gain' means simply that the measure [...] taps the form of thought, or the way in which reasoning is organized, rather than the specific content of the response, a distinction which Kohlberg's critics do not take into account." (Broughton 1978a, p. 3; see also Broughton 1978b; Lind 1989; Lind & Nowak 2015)

Authors who have proposed alternative methods of measuring moral development, often also bowed down to the dominant classical test theory. James Rest proposed measuring the development of moral competence by means of a questionnaire (the *Defining Issues Test*) which records a person's preference for principled moral reasoning. The hereby determined measured value *P* "indicates the extent to which a person considers principled arguments as the most important." (Rest & Narvaez 1991, p. 234) Other researchers have suggested similar measurement instruments (see Gibbs et al. 1992). Moral orientations actually often correlate astonishingly closely with moral competence because of an affective-cognitive parallelism (see page 74; Lind 2002). But this parallelism can only be shown in certain contexts. As simulation studies show, this parallelism can be dissolved as people can simulate their moral orientations in either direction if necessary (Emler et al. 1983; Emler et al. 1998; Haste 1985). Therefore, you can also easily influence the moral attitude test scores of learners with the help of lectures on ethics (Penn 1990). However, it is questionable whether the change of test values really indicates a change in the moral orientations of the learners, or – as in the experiment of Emler – only their ability to provide desirable answers (Lind 2002; see also here Section 4.5, page 74).

Wang et al. (2015) have suggested measuring morality and character through self-description questionnaires. For example, for assessing the virtue of 'honesty' (*trustworthiness*), they provide statements like "One can rely on me to tell the truth." In response, the participants are able to select on a 5-point scale, "That's not at all like me" to "That's me exactly." While it seems to some degree to be humanly possible that people can estimate their own moral competence it is very questionable whether they respond objectively to such questions. Wang and colleagues found no confirmation for their hypothesis that the character of young people is improved by their activities with the Scouts. They even found a decrease in test scores with age for non-Scouts. This strongly suggests, in my opinion, that their instrument does not measure morality, but rather participants' adjustment to the (alleged) expectations of

---

[7] The reader may ask why social scientists still cling to classical psychometric approaches, and why they use them in the first place. Psychometric tests of ability or competence originated mainly in the military in World War I and II, where they were created to select soldiers. That is, such tests have been primarily made to check whether or not people meet *external* requirements, not to study the nature of their abilities. Many test constructors subsequently made a career as academics in universities or in the test industry. They have a vested interest in applying their methodology, which may explain why the vast majority of psychological research uses classical psychometric approaches in spite of their obvious incompatibility with modern psychological research.

researchers; this adaptation usually decreases with age, only not in Scouts, who may be more obedient to authority.

In sum, classical psychometric approaches are not suited to assess moral competence:

• Tests which are based on classical test theory define and measure morality in terms of external standards – instead of the internal moral orientation.

• In order to display people's competence, a test must contain a difficult moral task. Otherwise the test scores can be simulated in every direction – as in moral orientation tests, but not in a true competence test. Competence can be simulated downward but never upward.

• Simply asking people about their moral competence overwhelms them. Firstly, they often have only a vague idea of their own moral competence. Secondly they are tempted to respond in a way which they think is expected from them. Thus, such questions measure more social desirability than moral competence.

Given this situation, I felt the need for a new measurement instrument.

4.3   The *Moral Competence Test* (MCT)

In search of an adequate moral task

Above all, a competence test must consist of appropriate tasks. Just as a test of mathematic competence must consist of mathematical tasks so a test of moral competence must contain moral tasks. If the test does not contain a difficult moral task, it cannot be considered a test of moral competence because it cannot differentiate between high and low competence of people. A sure proof of a competence test is the impossibility of faking its scores upwards intentionally.

What is a "difficult moral task"? Many tasks do not qualify because they represent external, social expectations. For example, letting another student copy one's test may be seen as "immoral" by the teacher, but from the student's perspective it may be meant as helping a good friend to make the grade. Other tasks do not qualify because their solution does not allow unambiguous inferences. For example, when we ask people how many crimes they have committed, their answers can hardly be considered a valid index of moral competence.

I believe that a better starting point for the search for a good test of moral competence is the definition of moral competence as the ability to think and discuss opposing views on moral issues. As philosophers like John Dewey, Karl-Otto Apel and Jürgen Habermas assert, moral discourse is the most basic moral principle. In the early 1970s, the social psychologist Charles B. Keasey showed in his experiments that most people are not able to discuss arguments that oppose their own stance on an issue (Keasey 1973). Isn't this also our experience in everyday life? If we discuss a controversial issue, things frequently become "personal" in a short time. We then no longer argue about the matter, but instead against one another, getting louder – or abruptly end the discussion by turning our backs on each other. The inability to weigh opposing views in our mind or to exchange arguments with opponents seems to be the main cause of deceit, violence and war. If we cannot settle controversies through peaceful, rational negotiations we can only "settle" them through immoral means.

Structure and content of the MCT

Against this background, the idea was born to measure people's moral competence by confronting them with arguments agreeing with and opposing their own position on a highly controversial issue, and by analyzing their pattern of responses. Would they judge the arguments merely on the grounds of opinion agreement, as most participants in Keasey's experiment did, or would they judge them consistently in regard to their moral principles regardless of the opinion agreement on the arguments? Below I will give a summary of the Moral Competence Test (MCT). More details on the MCT can be found in more technical publications (Lind 1978; 1985a; 2008b; for more sources see http://www.uni-konstanz.de/ag-moral/).

In the MCT, participants are faced with two dilemmas stories (vigilantism/workers, and mercy killing/physician), in which the protagonists make a decision. Participants evaluate this decision on a 6-point scale from right (+3) to wrong (-3). Subsequently, they are presented with several arguments for and against their evaluation. The participants are asked to say whether they would accept or reject these arguments. Actually they are asked to give a differentiated rating on a scale of -4 (completely reject) to +4 (completely agree).

Each argument contained in the test was selected or designed in such a way that each represents one particular type of six different types of moral orientation (see Section 3.2 above). For each moral orientation an argument in favor and against the protagonist's action is presented in the test. Because the MCT contains two dilemma stories, the participant has to rate 24 arguments altogether.[8]

I have taken many arguments from the interview examples in the evaluation manual for the *Moral Judgment Interview* (Colby, Kohlberg et al. 1987). When I did not find arguments my colleagues and I created new ones. Contrary to our expectations, we found it more difficult to formulate arguments of a low type than at a higher level. It took great effort to formulate arguments for and against mercy killing of Type 1 and 2 so that they do not sound ironic. That probably had something to do with our own moral feelings. How could one seriously discuss a case of mercy killing on the level of harm avoidance and monetary profit? We eventually succeeded in finding serious arguments for low type-reasoning by doing some role-playing.

All arguments were tested in two kinds of studies. To ensure that the arguments expressed as clearly as possible the type of moral orientation they were meant to express, we asked several distinguished experts on Kohlberg's theory to arrange the arguments according to the different types. On the basis of their reviews I revised the text of the arguments. To ensure theoretical validity of the MCT, we tested four criteria drawn from what we knew about the nature of moral competence (Lind 2008b): (1) participants should not be able to simulate their competence test scores upwards. (2) Participants should prefer "higher" types of moral orientation over lower types for discussing moral issues like mercy killing and vigilantism. (3) Participants' preferences for "neighboring" types of moral orientation (e.g. type 5 and 4) should correlate more highly than their preferences for more "distant" types of moral orientation (e.g., type 5 and 2). (4) Participants should prefer higher types of moral orientation the more clearly, and reject lower types of moral reasoning the more clearly, the higher their moral competence is. The first tests gave us hints for reformulating some arguments to improve their validity. After revising the MCT in 1977, many more validations have been carried out with

---

[8] The MCT is not a psychological test in the classical sense; rather it is an individual experiment in which hypotheses about the influence of certain dispositions on a person's judgment behavior can be tested. The MCT has a 2 x 2 x 6 factorial multivariate design, resulting in a total of 24 arguments (items). Through this design, we can test hypotheses over the influence of three factors on the participant's judgment behavior: the moral quality of the arguments, opinion conformity and situational context (Lind 1978; 1982; 2008b).

the original German version and with the 38 translations of the MCT in other languages. They all, without exception, confirm the four criteria almost perfectly (Lind 1978; 1985a; 2008b). A recent study by Don Biggs and Robert Colesante also shows that we were successful in keeping the grammatical complexity of all arguments constant. The ratings of the arguments, they found, did not correlate with the arguments' grammatical complexity (Biggs & Colesante 2015, p. 511).

The crux of the experimentally designed MCT is that the three factors of the test-situation are contained in each argument, rather than being operationalized in three different sub-tests. Because manifestations of each factor (six types, two sides of an issue, and two dilemmas) are present in each argument and are systematically varied and combined, it is possible to "read" the participants' minds from the pattern of the individual responses. When the pattern shows that an individual accepts all supporting argument, and rejects all opposing arguments regardless of the arguments' moral quality, her/his rating behavior has been solely determined by the will to confirm to a certain position. This way of dealing with arguments could be called, in psychoanalytic terminology, "rationalization". The subjects use arguments only to cement their position post hoc through moral arguments, rather than to examine their position critically, and possibly to revise it in order to align the position with their own moral principles. Conversely, when a pattern of ratings shows that an individual clearly rates the 24 arguments in regard to their moral quality rather than in regard to their opinion agreement, it is obvious that the individuals' moral principles have determined their rating behavior.

Six scales of moral orientation

The MCT also provides readings for each of the six moral orientations built into the test. Moral orientations are measured by averaging the ratings of the four arguments of each type of moral orientation (two for each dilemma story, one for and one against the decision of the protagonist). They are added together and divided by four. Each scale ranges (along the response scales) from -4 to +4. In some publications the sums of the scales are not divided by four, so that the measured values range from -16 to +16. The six orientation values are usually represented as a profile (for illustration see graph in Section 3.2, above). Older ways of indexing the affective aspect of moral behavior are not recommended anymore. The MCT Preference score (P-score), in particular, was phased out in an early stage of research because it is not a pure score of moral orientations. I had suggested using the MCT P-score analogously to the DIT P-score to measure how much a participant's preference order of the six moral orientations correlates with the hypothesized rank order of the six moral orientations. It was used, as far I know, only in two studies (Heidbrink 2010; Biggs & Colesante 2015). I dropped the MCT P-score long ago because its computation is confounded with the C-score. This means that the empirical correlation between the MCT C-score and the MCT P-score is partly spurious. It is as if we were to correlate the length of our legs with our total body length. Since both measurements involve the length of our legs, the correlation is partly spurious.

Scale of moral competence

The main purpose of the MCT is to provide a measure of moral competence (for its definition see p. 13). The MCT is, as you should keep in mind, a behavioral experiment, but not a classical psychometric test. Psychometric tests would be completely inappropriate because

they do not allow us to measure people's structures of moral judgment by internal moral standards.

The MCT produces what is called a "C-score", the "C" standing for competence. The C-score indicates to which degree a participant rates the argument of the test by their *moral quality*, rather than by other factors like their *opinion agreement*. In other words, the C-score is designed to show how able people are to engage in a moral discussion about a difficult issue rather than obstruct it by insisting on their opinion regardless of what speaks in its favor or against it. Such obstruction implies that the issue cannot be settled through discussion but only through other means like violence, deceit or the exercise of power. The C-score has been constructed so that it ranges from 0 to 100. A C-score of zero means that the individual's pattern of responses to the MCT does not manifest any moral competence. A C-score of one hundred means that the individual's response pattern perfectly meets the criterion of moral competence. The C-score is a numerical representation of the pattern of an individual's responses ("**x**") to the arguments presented in the MCT.

In the following graph, we depict the response pattern of two fictitious individuals in the order of the moral orientations which are expressed by the arguments. (In the test the arguments are presented in a random order.)

What do you see in the following graph? Think about this question for a moment before you continue to read my explanations.

Two Response Patterns Manifesting Different Degrees of Moral Competence (one story only)
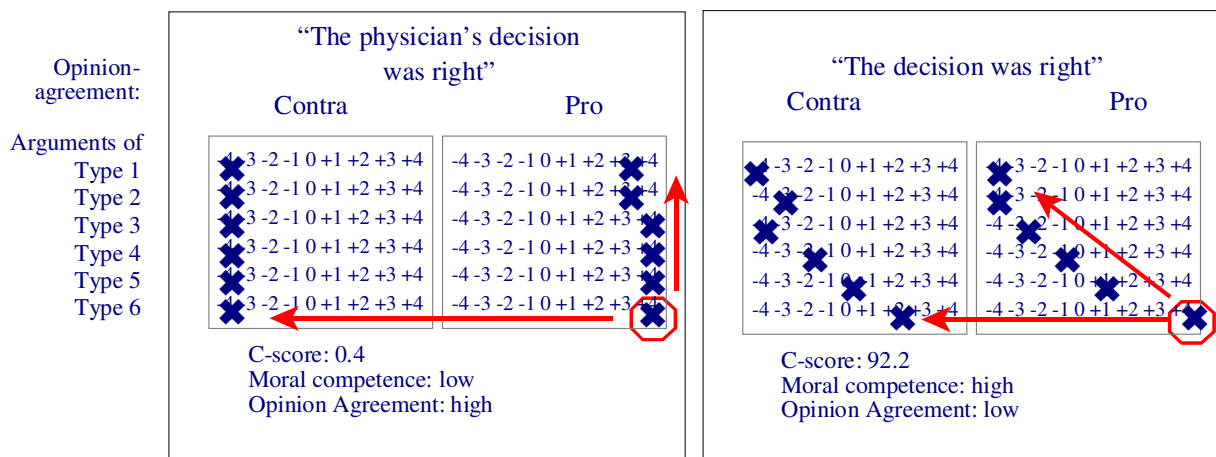
 Structural information can be assessed only from a *pattern* of individual responses to the test questions, not from isolated responses. Example: One very positive rating of a type 6 argument ("+4") signifies low moral competence in the left pattern, but high moral competence on the right side when seen in the context of the other responses. Note that in the MCT the arguments are not ordered according to their Type.

I hope you see the same things as I do (and possibly some more). First, I see that the C-score nicely reflects the different pattern of response of the two fictitious participants. The pattern on the left side shows almost no sign of moral competence and therefore produces a very low C-score. Obviously the participant rated the arguments only in regard to their agreement with his or her opinion on the protagonist's decision (here the physician in the second dilemma story), but did not let the different moral quality of the argument inform his or her rating of the arguments. On the right side we can see a response pattern which clearly signifies very high moral competence. It produces a C-score of nearly 100.

Second, you might have guessed why I have circled response ("+4") to the pro-argument of type 6. If we looked at responses to Type 6 arguments in an isolated way, as is required by classical approaches to psychological measurement, we would make a big mistake by interpreting this as a sign of Stage 6 moral development, at least in the case of the person on the left side. When being asked to rate a Type 6 argument opposing their stance on mercy killing, both subjects totally reject Type 6 reasoning. In the beginning of his research, Kohlberg used the technique of reversing the decision in his Moral Judgment Interview in order to test whether or not a Type 6 argument really signified a developmental Stage 6. Later the Kohlberg group unfortunately dropped this testing with counter-arguments but considered them as "non-chosen issue" and gave them less weight in the scoring (Colby et al. 1987, Vol. I, p. 161). Within the classical testing paradigm the two responses to Type 6 arguments would either be counted as a medium attitude toward principle morality (whatever this means), or the items would be replaced by other argument which better fit their statistical model – even if the new items had questionable theoretical validity. The structural information which we can see in the figure would be interpreted as "measurement error".

According our *structural* definition of moral competence, the calculation of the C-score is based on the entire *pattern* of a person's ratings. "Structural theory does not treat any changes as a change in structural competence unless the change is evident in a qualitatively new pattern of response. [...] A really new mode of response is one that is different in its form or organization, not simply in the element or the information it contains." (Kohlberg 1958, p. 498) As I already noted, unfortunately he and his associates did not follow up this notion but gave in to the criticism of mainstream psychometrics (Kurtines & Greif 1974).

According to our *internal* definition of moral competence the calculation of the C-score for a person is logically independent of the moral orientations of the person. What this means can be seen in the following graph of two (fictitious) individuals who have the same high C-score, but prefer totally different moral orientations when discussing a moral dilemma.
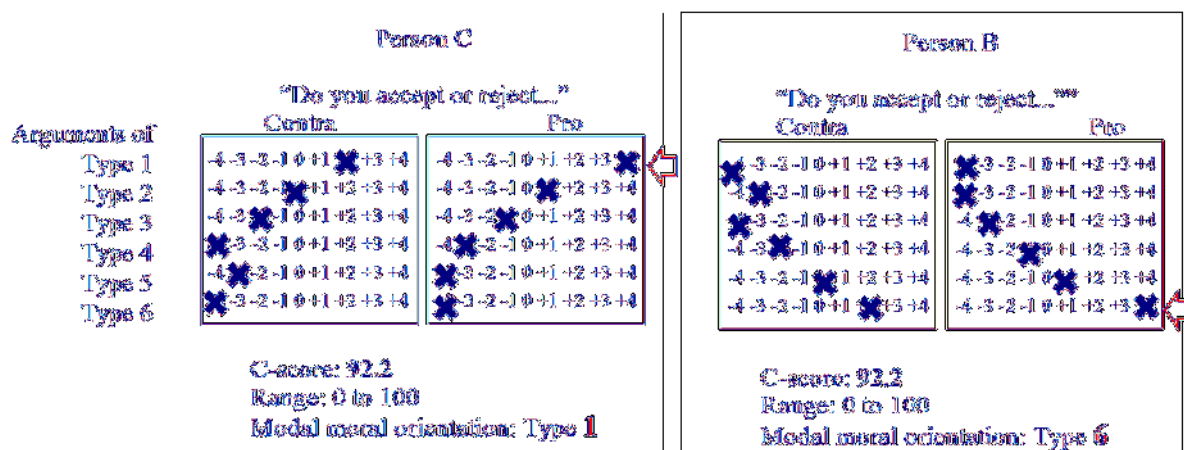
The MCT lets us measure moral competence of participants by their own internal moral standards. It shows two fictitious persons with the same high moral competence (C-score) and totally different preferences for moral orientations.

In order to find the C-score first one calculates how strongly the participant's 24 answers are distributed, that is, the size of the individual response variance (more technically speaking, the total sum of squares minus the mean sum of squares). Then one calculates how much of the variation goes back to the moral quality of the arguments, and divides this number by the total variation. The resulting proportion is multiplied by 100 so that a measured value is produced

which falls between 0 and 100. This measured value, the C-score, thus indicates how strongly a participant is oriented to the moral quality of the arguments, instead of conformity to his or her own opinion or other factors (Lind 1978; 1985a). Average C-scores hardly ever exceed 40.0; higher values are rare, which is another point supporting our claim that completing the MCT is a difficult task for many people. Therefore it has become common to graphically represent the C-score on a scale from 0 to 40, except when values higher than 40 occur. This uniform representation allows for easier comparison of the results of different studies with each other.

The C-score is always measured in the same way in each study and is therefore comparable, in principle, across all trials. This also applies to international studies since translations of the MCT must always be subjected to strict and elaborate validity testing before they are certified as valid (Lind 1986a; 2008b). Strict requirements must be met to ensure that the values are equal in meaning with the original test. It is not enough that the arguments are translated correctly. We know the problems that arise when we speak a foreign language. Even if we know the language well, we still have issues with the ambiguity of terms. A translation may be entirely correct, and yet may have very different meanings for our participants than what we intended. The validity studies can specifically detect ambiguity in translation, even if you do not speak the language. On the basis of these validity studies most translators immediately recognize that an argument needs to be changed so that it is equivalent in meaning to the original. Before such changes are accepted, they must be checked again in another validation study.

I am frequently asked whether one can shorten the standard MCT by dropping one of the dilemma-stories, or substitute the stories by other stories which may have more "face-validity" for a particular research question or for a particular population. Obviously, the standard MCT is a structural whole; if one breaks parts of it away, the C-score cannot be calculated anymore in the way we have described it above. Of course, one can also calculate the C-score for a single dilemma story – and this is actually done in so-called moral segmentation studies (Senger 2010; Bataglia & Schillinger 2013) – but these scores cannot be compared directly with findings from studies using the standard MCT. It is a common mistake to believe that the measurement of a psychological concept should or can be adapted to a particular research question or population. Imagine if we were to exchange our yard stick to align it with our objects. No comparisons could be ever made, and no scientific knowledge gained. This does not preclude critical reflections on the suitability of the standard MCT for one's research intentions and one's research subjects, or the need for new dilemma stories. I have been involved in some attempts to construct a new story. Only one has succeeded so far, namely the attempt by the Brazilian psychologist Patricia Bataglia (Bataglia et al. 2007; Bataglia 2009). The new story is best used as an addition in research studies, not as a substitute for the standard MCT.

Application of the Moral Competence Test

The MCT can usually be used with people over the age of about ten years. The test requires basic reading skills. If some words are not understood, the supervisors are allowed to help as long as they do not influence the participants. Our many years of experience and studies with the MCT demonstrate that participants usually fill it out gladly, because, as some say, the stories are very interesting. Questions are rarely skipped. It seems that this happens mostly by accident. In the electronic version, you can remind the participants to check their answers for completeness before they are sent in.

The MCT can be completed as a group test (e.g. in class at school), as a single test filled-out by individuals, sent by postal mail, or filled out online. To make sure that the participants fill it out carefully, some provisions should be made. Participants are more motivated to fill out the test carefully if they feel that they are participating in an important study, for example when they are asked to help to evaluate the effectiveness of a moral education program rather than producing data for a thesis research. Researchers should always report the return rate in their publications because it can have an impact on the C-score. If many participants drop-out, the C-score of the study group can increase because the drop-out rate is correlated with low moral competence (Krebs & Rosenwald 1977). Filling in the standard version with two dilemmas lasts between eight and fifteen minutes depending on the reading speed of the participants. No time limits are specified for the MCT, but in the case of a lengthy consideration of the answers encouragement for faster completion can be gently given.

Ability tests such as the MCT can react sensitively to specific external circumstances of the test application. The following circumstances can lead to a substantial reduction of the C-score:

- Time pressure: the MCT should be completed without a time limit, as already mentioned. Only when participants require excessive amounts of time are they encouraged to take less time.

- Test taking fatigue and priming: The C-score is lowered when the participant gets tired and also when the MCT follows to an achievement test. Preceding achievement tests prime him/her to think that the MCT requires to meet some social expectations rather than their own judgment, Therefore, the MCT should always be completed as close to the beginning as possible.

- External expectations: the C-score usually decreases if the participant gets the impression this is an audit of their "performance". In one study, this impression was primed by an intelligence test which the respondents had to fill out before the MCT. The results were much lower than should be expected.

Comparisons across different conditions of test-taking easily can lead to wrong conclusions, especially when the authors of the studies fail to draw attention to these circumstances. However, comparisons made within the same study are usually not affected. Generally, good knowledge of moral psychology is necessary for the application, evaluation, and interpretation of the MCT. (Lind 2008b)

Although the MCT measures individual moral competence, it is not suited for individual diagnosis or selection. We do not know whether it measures precisely enough for this purpose. In contrast in the research and impact or efficacy studies, typical values are averaged from multiple test subjects (there should be at least fifteen) so that the result is usually accurate enough to interpret differences of a C-score.

The MCT is not allowed to be used for individual diagnosis and selection because the test could be corrupted very quickly; as is the case with other selection tests, a lot of time and energy would be spent to cheat in completing the MCT. After a while, the test would then measure the corruption efforts of the participants, and no longer their moral competence. It could no longer be used for serious research. After all, the MCT has already been used for forty years, without any need to revise or replace it. Hardly any other psychological test has been used unaltered for research and curriculum development for such a long time, or has produced a comparable wealth of findings.

Because the MCT is based on a new method of measurement, it invites misrepresentations. Because the participant has to "recognize" the questions and arguments which are to be rated, some call it a "recognition test". However, they overlook the fact that the MCT does not measure how well participants can recognize individual items (as is the case in classical test theory), but measures moral competence which cannot be merely recognized. The C-score is often translated into a "consistency score". This wrong interpretation is also influenced by the prevailing test theories. There it means how great the variance of people's overall scores is compared to the variance between respondents and the variance. There are three misunderstandings involved here: (a) as psychologists we should be interested in consistency of an *individual*'s response pattern, and not in the consistency of an accidental sample of people; (b) there is no "consistency per se" (there are many kinds of consistency) but the term consistency is meaningful only if it is defined in reference to a specific criterion. So one could rightly say that the MCT measures individuals' *consistency of ratings in regard to their individual moral standards*; (c) the C actually stands for <u>c</u>ompetence.

4.4 How important is moral competence for "behavior"?

Before we can answer this frequently asked question we must clarify its meaning. There is an ambiguity with the term "*behavior*" which does not only lead to problems in everyday life, but also in moral research and moral education. There is hardly any term which looks simpler and, at the same time, is more ambiguous. This ambiguity has its roots in two wrong assumptions. First, most overt human behavior is not the result of a hard-wired reflex arc, but rather the result of complex cognitive processes, even those which are called affective, or emotional reaction to threat. Second, behavior can mean something completely different from an internal point of view than from an external point of view. Therefore the question how moral competence and "behavior" are related requires careful consideration: when we observe a certain behavior, do we really observe the behavior which we mean to observe? Already half a century ago, the renowned psychologist Donald T. Campbell discovered that behavior as simple as the behavior of rats which take part in a maze experiment can be easily misunderstood because of a less-than-perfect "conceptual overlap" between the experimenter and the rat. How much more complex is human behavior! Campbell concluded that "in no case should a single overt behavior be regarded as the criterion of a disposition." (Campbell 1963 p. 162)

As we have seen in connection with the measurement of moral competence (Section 4.2), behavior is usually the result of a complex decision-making and assessment process, which does not need to run consciously and which we therefore are not fully aware of or clear about ourselves. Behavior often means something different depending on whether we ourselves are the doer, or whether we observe the behavior of another. Others might not understand the intentions that we associate with our own behavior. Accordingly, the criteria by which we measure moral behavior could be quite different if we measure it based on "internal standards", that is, on the moral orientations of the individual, or "external standards", that is, on the moral expectations of others. These can overlap but are not necessarily the same (see graph on page 41). I think we would stand to gain a lot if, in considering the content relevance of moral competence, we differentiated between the inner and outer perspective, that is, between the behavior as it appears from the perspective of the actor and from the perspective of others (including observers and educators).

You may wonder why I do not also examine the question of the impact of moral orientations (attitudes, values, etc.) on behavior. Well, as I argued above, orientations are undoubtedly

very important for moral behavior as there would be no moral behavior without them. But they cannot correlate with differences in moral behavior because all people basically – apart from slight variations – have the same moral orientations (see Chapter 3).

4.5 How moral competence is manifested in behavior (internal perspective)

Is there really such a thing as moral competence? Can it be shown as a manifest disposition in human behavior? (Another question is whether moral competence has any impact on social behavior, i.e. the behavior that others expect from us. I will deal with this issue in more detail in Section 4.6).

The question of whether the judgment of human behavior is at all influenced by moral orientations can be formulated more precisely. Eminent moral psychologists have formulated some specific hypotheses for this assumption:

- Based on Socrates' observations, we assume that the ability to do the good (he called it "virtue," we call it moral competence) is broadly distributed in contrast to the will to do good, that is, in our terminology, to the moral orientations. In fact, a very broad spread of C-scores, which is the indicator for moral competence, is evident in all trials of the *Moral Competence Test* (MCT, see above, Chapter 4).

- All people have high moral orientations (see above, Chapter 3). Even criminals have high moral ideals. (Levy-Suhl 1912; Nucci 1995, p. 25; Hemmerling 2014) Hence violence, fraud, corruption and abuse of power in the world cannot be explained by a lack of moral values or principles.

- Jean Piaget assumed that there is an empirical parallelism between affective and cognitive development. (Piaget 1976; 1981) He did not test this assumption, probably because he lacked appropriate measures. Kohlberg and his associates studied this relationship using the *Moral Judgment Interview* to measure the affective aspect, and a Piagetian logic test to measure the cognitive aspect. (Kuhn et al. 1977) They found some evidence for a time-lagged development, the logical abilities being a trailblazer for moral development. However their study had a severe drawback. They dealt with both aspects as if they were separately measurable components and as if the MJI was a pure measure of moral affect, and not of moral cognition. Because the MCT allows us to measure both aspects as distinct but inseparable aspects, Piaget's assumption can be tested for the first time: is there a close empirical relationship between moral competence on the one hand and moral orientations on the other? Do people prefer high-type moral orientations and reject low-type orientations, the higher their moral competence is? As the following graph shows, our findings from a study with more than 2000 university students confirm Piaget's prediction with unusual precision. This finding has been reproduced in all subsequent studies with the MCT in many countries. (Lind 1986b; 2008b) The failure of Biggs and Colesante to confirm the parallelism assumption in their study seems to be due to their selection of measures: while one (MCT C-sore) is a pure measure of the cognitive aspect, the other (MCT P-score) confounds both aspects. Hence the correlation between both measures is partly a logical tautology producing a curvilinear artifact.
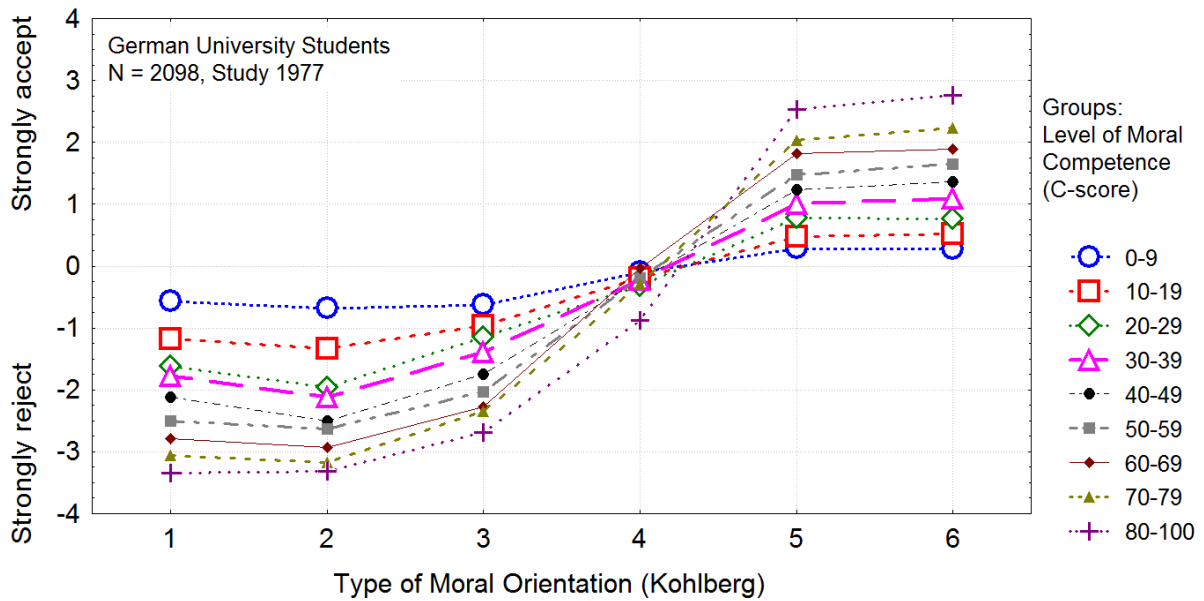
Empirical proof of affective-cognitive parallelism: the higher the moral competence of the participants the more they accept high-type moral orientations and reject low-type orientations when discussing a decision in a moral dilemma-story (Lind 2008b).

- If morality at its core is a skill, it cannot be simulated upward (see below) but can be increased only through (good) education. In fact, of all the variables studied, the length and quality of people's education correlate the strongest with the level of their moral competence (Lind 2002; Rest & Thoma 1985; Rest & Narvaez 1991). Most correlations with other variables such as age, social class, gender, and culture are much lower or very small. If there is some correlation it mostly disappears once the level of education is held constant, which accounts for the actual difference (Lind 2002; 2015a; see also Chapter 5). High correlations between moral competence and chronological age can be exhaustively explained through the level of education, which, of course also correlates with age. If there is age-related correlation left even after education has been partialled out, it seems to be an artifact of certain instruments which, like the MJI, have been optimized for the correlation with age (Colby et al. 1987; Lind 1989). However, it is the quality, not the quantity of education that correlates with the level of moral competence (Schillinger 2006; Saeidi-Parvaneh 2011). Some schools and universities do not promote more competence. In some countries no correlation exists between moral competence and level of education; most institutions of education seem to fail to promote moral competence. (Feitosa et al. 2013; Kukolja 2014; Liaquat 2013; Lind 1986a; 2002; 2015a; Lupu 2013; Saeidi-Parvaneh 2011; Schillinger 2006; Slováčková & Slovacek 2007)

- Finally, considering the nature of morality as ability shows that it cannot be simply simulated upward as is possible with moral attitudes. Competencies can only be simulated downwards.

The last assumption has been proved experimentally. I would like to elaborate on this point because it is central to the concept of moral competence. The idea of an experimental apparatus to answer this question comes from Nick Emler and his colleagues, who wanted to test the central assumption of the competence theory of morality (Emler et al. 1983). Their experimental setup was as follows: the test subjects, students, were firstly divided into three groups according to their own statements based on their self-rating as politically left-wing/liberal, neutral, and right-wing/conservative. Then the experimenters presented them with a test of preference for principle moral reasoning, Rest's already mentioned *Defining*

Issues Test (DIT). The DIT contains a series of dilemma stories like the MCT. But unlike in the MCT, the respondents are not presented with arguments for and against their own opinion, but instead with statements ("Issue Statement") thematizing the various Kohlberg-Stages. These should be rated by the subjects according to their importance and then ranked. Based on the answers, a so-called P-score is then calculated for the participants, which can vary from 0 to 95, indicating how strongly they prefer principled moral orientations over lower type moral orientations (Rest 1979; Rest et al. 1999).

In its first part, Emler's experiment showed that the DIT P-score was highly correlated with the political attitude of the subjects: left-wing participants received significantly higher P-scores than right-wingers (see also Gielen 1986). Due to this finding, Emler and colleagues confirmed their suspicion that morality has to do with a disposition, similar to a political orientation. They undertook a second experiment in order to further show that morality is truly an orientation or attitude (but it is not a capability). The same subjects were asked to put themselves in the position of the politically opposite set of people and to complete the DIT as they imagined these people would respond. If it were possible for the test subjects to simulate their judgments in any desired direction, then this is proof that morality cannot involve the expression of a capability. Of course, particularly critical would be the simulation of test scores that are higher than one's own, which means here simulating the moral rating of the left-winger by the right-wingers. The illustration below depicts the result of an experiment by Emler and his colleagues. It shows that it is indeed possible to simulate the DIT P-index upwards. The group of self-described right-wing students who scored low in the first round can now improve their P-score (straight line) to the level of the left-wing students (dashed line) when asked to do so. The fact that the left participants could just as well simulate their values downwards is not surprising. Emler et al. (1983) proposed two conclusions from this result: first, there is no moral competence, and second the DIT is not suitable for measuring such skills. While the authors seems to prefer the first conclusion, I think the second conclusion is more likely because the DIT is a measure of moral preference, but not of moral competence.

My colleagues and I re-staged the experiment with the MCT, which – unlike the DIT – contains a moral task and measures moral competence. Here, too, students were examined and split into groups according to their self-reported political stance They were then asked to perform twice, the first time to determine their own test scores and the second time with the instruction to complete the test from the perspective of their political opponents (for further details see Lind 2002). The results of our experiment are displayed in the graph on the graph below. In agreement with Emler's experiment, students describing themselves as right-wingers had a lower moral competence (straight line) than students who rated themselves as left-wingers (dashed line). In contrast to the Emler experiment with the DIT, the subjects failed to simulate the test scores upward. They were unable to obtain a higher C-score under the simulation condition than they showed under the regular condition. Participants' MCT C-scores in the simulation condition were always below those they themselves had previously reached. Wolfgang Wasel reconfirmed our findings in a modified experiment using the MCT. His study showed, moreover, that the accuracy of the simulation is highly correlated with the level of the subject's own moral competence: someone with high moral competence can better recognize this ability in another than someone with lower moral competence (Wasel 1993). Both experiments prove (a) that there is such a thing as moral competence and (b) that it can be measured with the MCT.

| Moral orientation The Emler-experiment (DIT) | Moral competence The Lind-experiment (MCT) |
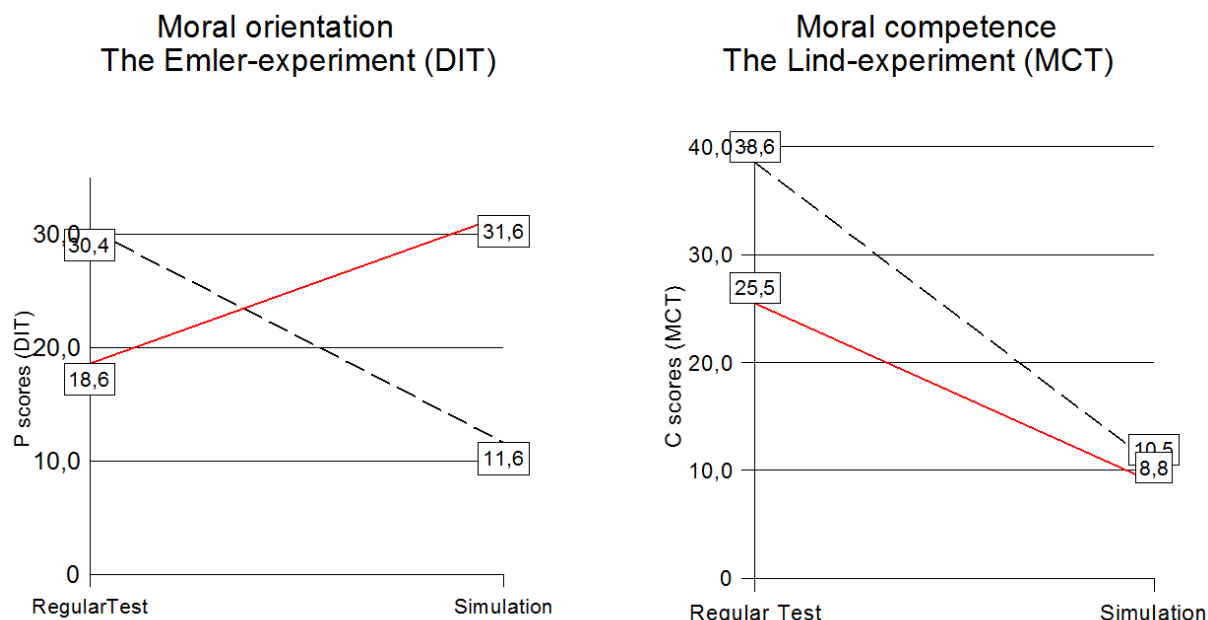
Left graph: moral orientations (preference for principled reasoning) can be simulated upward. Right graph: but moral competence cannot (see text). Left-wingers' scores are connected with a dashed line (- - -), right-wingers' score with a straight line (——).

4.6 How moral competence impacts social behavior (external perspective)

We would expect a very high correlation between people's moral competence and the behavior that is expected by society. But this expectation would only be justified if we lived in an ideal democracy, that is, if there was a complete overlap between peoples' shared moral principles and society's rules and laws (see the graph in Chapter 2, page 41). People would transgress rules only because of lack of moral competence but not because they oppose them.

Most existing democracies are far from being perfect. Therefore, low correlations do not always point to a lack of moral competence, but can also point to dubious rules and laws. There is, to my knowledge, no research that provides data for testing this hypothesis. But this thought experiment shows that the correlation between a moral competence test and externally rated "behavior" would be a dubious indicator of test validity. Therefore we did not rely on "predictive validity" for validating the MCT. Instead we used criteria derived from grounded theory about the nature of moral orientation and competence (Lind 2008b).

A second reason why we should interpret correlations between moral competence and "behavior" with caution is that we know from research as well as from everyday life that the observation of behavior is hampered by a number of difficulties: we know that people's behavior is often influenced by the conditions of observation, e.g., by the gender or the social status of the observer, by the criteria applied to the observation (when is taking something away to be called theft or robbery?) and by the diligence of the observer (most criminal statistics reflect as much the criminality of a country as the eagerness of its police to report particular data). Moreover, as we have seen, the correlations obtained from such observation studies depend very much on the distribution of both sides of the equation. If either the distribution of moral competence or of social behavior is close to zero there will be low correlations for purely mathematical reasons. Even worse, in case of a curvolinear correlation in a population, the study of sub-groups whose C-scores are restricted to the lower or to the higher range of the scale can yield completely opposite correlations. Many researchers do not give any information which would let the readers come to their own conclusions.

How to teach morality

The best source for valid data on the relationship between individual moral competence and externally rated social behavior is provided by carefully planned experiments. Contrary to common belief such experiments exist. They examine question like these: do people with high moral competence display more helping behavior (and not merely willingness to help)? Do they show more rapid decision-making behavior? Are they less violent and display less criminality? Are they mentally healthy? Can they more effectively oppose immoral orders? Do they tend less toward stereotyping and pluralistic ignorance? Do they more strongly advocate democratic liberties? And does moral competence perhaps affect learning behavior?

The question how moral competence relates to "social behavior" raises some issues concerning the very meaning of social behavior that we should consider before we look at these experiments in detail:

- Like moral behavior, social behavior cannot be determined from the observation of a single act or decision (Simmel 1989, p. 93; Brunswik 1955; Lind 2008b). When we observe a single act, we are only provided information about its compliance with certain social expectations. In order to determine what is really indicated by "helping behavior" or "honesty" we must know more about the circumstances and context of this behavior. We must be able to exclude the possibility that the behavior means something different than what we think it does. The use of so-called hard data such as public statistics and files often does not help because they are not always reliable. Often there are reasons to exaggerate or to cover-up certain problems such as theft and drug use. In other words, if we want to observe social behavior, we need to very carefully plan the observation of the behavior and conduct, which is usually associated with a lot of effort and high expenses. But superficial observations can lead to wrong conclusions – and hence to wrong educational policies.

- If we desire to study the impact of moral competence on social behavior, we have to actually *observe* behavior: helping behavior, honest behavior, resistance to immoral suggestions, etc. Therefore, we must not be content with asking people about their behavioral intentions or to have them report how they would behave in certain situations. Intentions are often – as I am sure most of us know – not implemented for various reasons. For example, when our motivation is weak or because circumstances change readiness to lend help does not really signify helping behavior. Furthermore, self-reports on our behavior are not a reliable substitute for the direct observation of behavior. Self-reports are often colored by the desire to appear positive. It is also possible that people understand our questions or requirements differently, for example in regard to the use of terms like "often" and "sometimes."

- When we want to study the effect of moral competence on a certain behavior, we must not rely solely on correlation studies; we must try to answer our questions with carefully designed intervention experiments.

- Finally, the classification of behavior as "helping", "cheating" or "violent" for example, is not always clear from the social point of view. This classification often depends on the passage of time and the standpoint of the observer. What was considered criminal fifty years ago could now be considered normal and vice versa. Homosexuality and sexual relations with minors in my country are examples. Official statistics of "deviant" or "criminal" behavior cannot therefore be taken without basic testing as an indicator for the norm-conforming behavior we want to study. In our experience, some of the schools we surveyed very willingly provided information regarding rule violations, while others preferred not to as they were afraid of earning a bad reputation. Often students' reports come closer to the truth.

To answer the question of whether moral competence has an impact on social behavior, we have a number of carefully-scale experimental studies on hand. I limit my presentation to those studies in which moral competence was measured, i.e., studies that were carried out with the MCT (Lind 2008a; Lind & Wakenhut 2010) and with the *Moral Judgment Interview* (MJI) of Kohlberg and his collaborators. (Colby et al. 1987) As I have demonstrated in Chapter 3, the MJI has some drawbacks, especially since it provides a confounded measure of moral competence and orientation. Yet at least it does tap the moral competence of participants and not only their moral orientations.

I also restrict my presentation – with few exceptions – to those experiments in which *actual behavior* was observed and not merely behavioral intentions or self-reported behavior. Unfortunately, many authors do not distinguish between these forms of behavior, and too often they blur the difference. They do not seem to realize, for example, that actually helping persons in distress is something quite different from the intention to help, and saying that one would help. All studies which I summarize below have already been described in detail elsewhere in more detail. (Kohlberg 1984; Lind 2002; Sprinthall et al. 1994)

Helping behavior

Is moral competence an important factor for helping behavior? Sharon McNamee (1977) has done one of the very few experimental studies on this question. She asked subjects to participate in a laboratory experiment on the subject of learning. The actual trial had already begun, however, during the time participants were waiting to be called for their turn. Suddenly, the door to the lab opens and a person comes out and collapses in front of the waiting room. The person appears lifeless. Meanwhile, the experimenter observes the waiting room unnoticed, counting how many respond to the incident. The participants react differently. Some stand up and try to help, while others sit idly by. Previously, on another occasion, the same subjects were surveyed using the Kohlberg interview so that there was no recognizable connection with the helping behavior experiment.

The analysis of the observations showed a very strong correlation between the "moral Stage" of the test subjects and their behavior in the experiment. While only about 20 percent of the subjects tried to help who had Kohlberg's Stage 2 on moral judgment, this proportion increased sequentially from one Stage to the next. More than 80 percent offered their help at Stages 5 and 6. This finding confirms the thesis that moral competence is very important for people's pro-social behavior.

Additional light was shed by further analyses and inquiries. Firstly, it appears that (with the exception of Stage 2 participants) all had a very high degree of helpfulness, (i.e., that their intention to help was great. There was, therefore, a gap between the intent and the execution of the aid response for many individuals. This gap became smaller with increases in moral competence. Why did participants with low moral competence not assist, although they expressed similar willingness to help as those participants with high moral competence? Subsequent enquiries showed that both groups had similar thoughts when they were confronted with the (feigned) collapse: they saw a need to help and wanted to help in most cases. They also found themselves confronted with different ideas for and against helping in this situation, wondering, "Why did the man collapse?" "Is he sick or – even – dead?" "What can I do?" "Can someone else help better?" "What happens if I do something wrong?" The participants with high moral competence, however, were faster to clarify the conflicting considerations and to reach a decision than the low moral competence group who sat motionless.

People who do not help when it is necessary, therefore, are no less moral than others. They have, as these and other studies show, often the same high moral ideals and are just as willing to help. They lack only the ability to make decisions in a short time. In many situations this is crucial.

Decision-making

The experiment of Franz-Josef Mansbart (2001) provides further insights into to the above interpretation. In this study, participants had to make decisions on moral dilemmas that were presented to them on a computer as fast as possible. Mansbart also assessed the moral competence of the participants using the MCT and gathered information about their motivation to make quick decisions in such situations. The findings showed that the rapidity of decision-making (in seconds) in this situation depended much more on the degree of participants' moral competence development than the strength of their motivation to make a decision in this situation. The relative effect size was ten times larger for moral competence ($r = .36$) than for the four motivation variables included in Mansbart's experiment ($r = 0.03$). Of course, the speed of a decision is not always a good indication of ability to make decisions; sometimes people feel compelled to make a decision before they are able to obtain and evaluate all the necessary information. In Mansbart's experiment there was indeed a time constraint, but no social pressure associated with sanctions. I do not doubt that the subjects gave their verdict only when they felt certain about their decision.
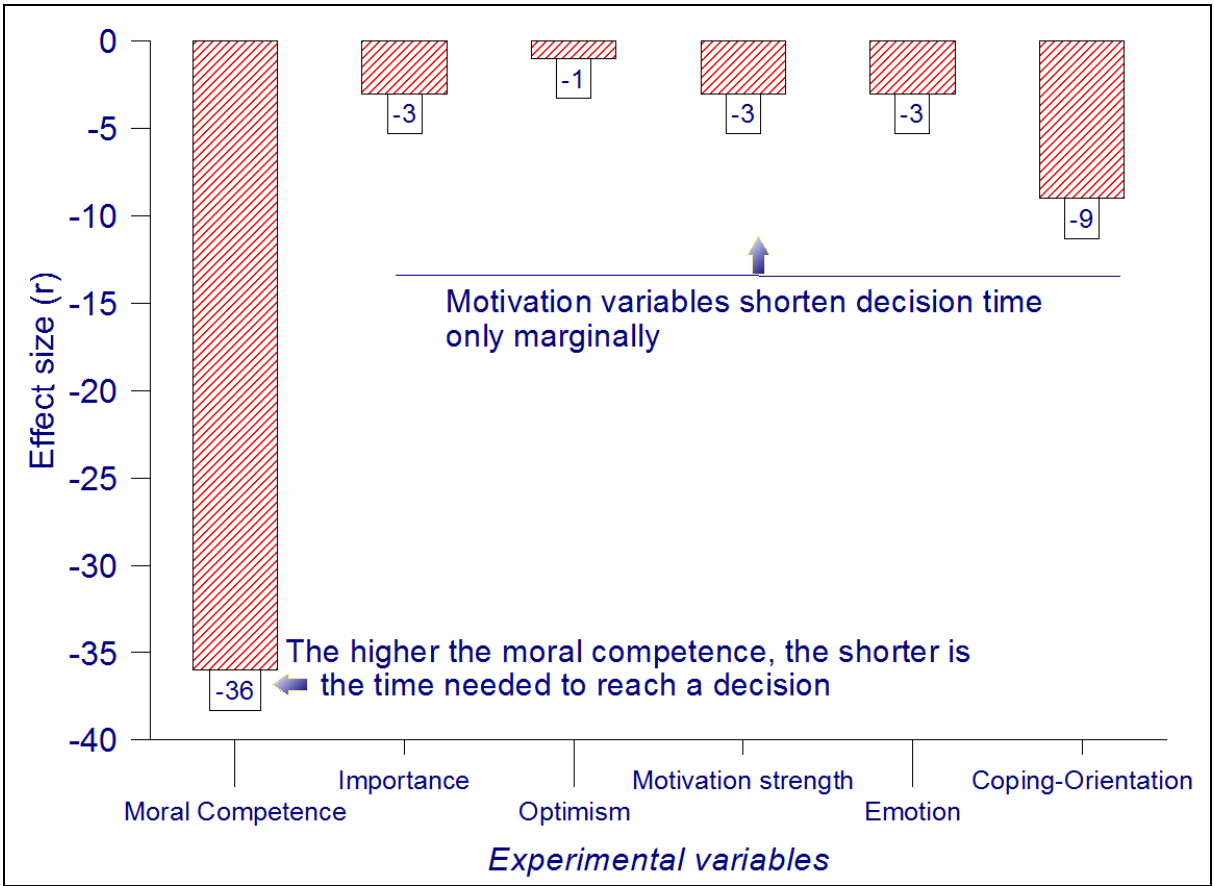
Influence of moral competence and various motivation variables on decision time: the higher people's moral competence the shorter time it takes them to make a decision. Source: Mansbart (2001); my calculation of effect-sizes (GL).

From these experiments we can then conclude that people make morally good decisions when they: (a) have the ability to solve conflicting moral requirements (dilemmas) and (b) have the required time. Thus the quality of a decision seems to depend more upon well developed moral competence than on motivation. Probably additional participants in McNamee's experiment would have offered their help if they had had more time to do so. The experiment was nonetheless close to reality, as we are often confronted with situations in which decisions must be made quickly because otherwise the aid may come too late.

Kristin Prehn and her colleagues did another revealing neurological experiment at the Free University of Berlin. (Prehn et al. 2008; Prehn 2013) They presented participants with brief statements for assessment which contained either a moral or a grammatical problem: "He breaks a window" (moral) and "He *look* out the window" (grammatical). These were then compared with similar statements that contained no problem: "He look̲s out the window." The subjects had to quickly decide whether the statement contained a problem by pressing a button. The researchers measured the error rate and the reaction time as indicators of the ability to make decisions. An additional aspect of the experimental setup is important to us here. During this experiment, the subjects were in a brain scanner, which recorded the brain regions these task activities triggered and their intensity. The intensity of these activities was indicative of how difficult it was for subjects to solve the task. Brain activity in the right dorsolateral prefrontal cortex (DLPC) was particularly strong if they were moral tasks. The fact that this region of the brain is particularly important for moral competence is demonstrated by the activity level's (BOLD-level) high (negative) correlation with the subjects' MCT C-score: $r = -0.47$. The higher the moral competence of people, the less effort they need to make moral decisions. That is, this study also affirms the hypothesis that moral competence is a very important factor in human action. We can easily imagine that people with low moral competence and, therefore, with a need for much time to make a decision, will experience problems and conflicts as being much harder and feel more painful emotions than people who can make good decisions more swiftly. A further brain scan study by Kirstin Prehn and her colleagues supports the notion that moral competence scores reflect people's ability to handle emotional conflicts when faced with moral conflicts (Jung et al. submitted for publication).

Refraining from the use of drugs

What does moral competence have to do with drugs? Some consumers use drugs to create pleasant feelings, but some also use them to control emotions which are perceived as painful and can severely restrict one's quality of life. Such painful emotions often stem from dilemmas that take much time and energy to solve or turn out to be insoluble for the individual.

When we face a decisional conflict (dilemma), we first notice our feelings. We usually feel a dilemma before we perceive it consciously. We can trace the conflict between opposing moral feelings in a more rapid pulse, higher blood pressure and rumbling in the belly. Often we resolve such dilemmas purely emotionally, that is quickly and without a whole lot of thought. We achieve this more easily, as I have explained above, the better developed or trained our moral competence is. Feeling and resolution of a dilemma often happens so fast that we hardly notice it consciously. If this kind of automatic response does not occur, we may notice that we are activated or stressed just as if we were threatened by some danger. A problem whose resolution is urgent is indeed much like a threat. We can hardly think of anything else apart from how we can solve the problem: "Why is my friend not speaking to me? Is he angry with me? Should I apologize? Or should he not take the first step?"

If the problem is very difficult, it can take a long time until we can solve it, and our stress remains. If this state of excitement stays at a high and painful level, it can make us sick: we

can hardly sleep, concentrate on our work or education, and otherwise act oddly. Proper treatment of this state sometimes requires more expertise, time and money than is available to us. Then we often endure this state until a solution is found or we try to numb the excitation by consuming alcohol, cigarettes or marijuana. These substances are relatively inexpensive and readily available. Drugs are a help but they are certainly not a real help. They may be used as a kind of "self-therapy" but it is one that tends to lead to new problems in the end instead of solving the original problems.

The study by Beke Lenz among youth confirms this theory. She shows that in general only a weak correlation exists between moral competence and drug use, but that this relationship is strong when one compares young people who have many life-problems – such as difficulties at school, parents' divorce, loss of friends, and so on. Within the group of youth with many life-problems those with low moral competence show higher consumption of alcohol, cigarettes and marijuana than those with high moral competence. (Lenz 2006) It seems that young people with higher moral competence can solve their problems and conflicts by thinking and talking, and thereby do not need to consume drugs in order to skirt painful emotions. Moral competence then seems to lower the risk of becoming a drug addict when confronted with blows of fate or just with difficult situations at school or work. While we can hardly prevent the latter, we can foster youth's moral competence in order to protect them against harmful "self-therapies".

Following rules

"Deviant behavior is not an illness that has attacked the individual but is an expression of the total person as he actively, although ineffectively, attempts to come to terms with his environment. Persons who behave inappropriately need help in order to develop means for effective participation in the life of the community," writes Leslie Phillips (1967, p. 232). Most people, according to our thesis, do not resort to means of violence and breach of the law because they lack moral values or even have anti-moral values, but because they lack the ability to recognize their own moral principles and to translate them into appropriate behavior. We are convinced with Socrates that whoever understands what is good has no choice but to do the good. Violence, fraud, corruption and war, it seems, are types of conflict resolution, but surely the lowest among Kohlberg's six types of moral orientations (see page 53, and also Lind 2010d).

Evidence for a relationship between moral competence and criminal behavior (violence, fraud, corruption, abuse of power) is found in many in correlation studies. The lower the level of people's moral competence is the greater the risk of their coming into conflict with the law. (Blasi 1980; Hemmerling 2014; Kohlberg 1958; Weyers 2005) The relationship is somewhat stronger with criminal behavior involving the use of violence and brute force than with white collar crime. (Wischka 1982) Yet also the role of low moral competence in the latter is well documented in correlation studies. Stemming oneself against corruption and illegal practices in firms and organizations by blowing the whistle and showing the red flag seems to require high moral competence. (Brabeck 1984; Roberts & Koeplin 2002)

This relationship between low moral competence and criminal behavior is even more clearly demonstrated by experimental studies in which – as in the studies of Hartshorne and May (1928) – subjects were tempted to cheat in class tests or exams. The lower the moral competence of participants was, the greater the proportion of those who cheated. (Sprinthall et al. 1994; Kohlberg 1984) When we tested the method of dilemma discussion and the Just Community (see below, Section 11.3), the evaluation showed that the intervention not only

increased the level of moral competence of students, but also decreased the number of infractions. (Lind 2002; Lind & Althof 1992) Kay Hemmerling (2014) likewise reported that the number of violations in a prison department were generally lower due to the KMDD sessions held in the prison.

These findings refute the opinion – to be found even in textbooks on criminology – that moral competence has no relevance for criminal behavior. Contrary to this belief, we can very effectively resocialize the accused by supporting their moral competence with the KMDD. The application of this method in schools also contributes to the prevention of violence, fraud, and abuse of power. (Lind 2002; Hemmerling et al. 2009; Hemmerling 2014)

In the applied project *Democracy and Education in Schools* (see Section 10.3) there were also indications that the promotion of moral competence through dilemma discussions and *Just Community* meetings decreased the number of rule infractions such as bullying, vandalism, theft, and truancy by pupils. (Lind 2002) A new study of German children ranging from 11 to 17 showed a marked correlation between low moral competence and reported school bullying. (Grundherr et al. 2016)

Obedience to authority

Many social institutions function only if orders from one authority or another are followed. The person who gives the orders also takes on the accompanying responsibility, while the persons who carry out the order are responsible for their own actions even though they are following directions from above. The underlying democratic idea includes recognition that a superior who orders an act can never assume full responsibility for it as orders can never be perfect (as is also true of laws and regulations). Circumstances may arise that management cannot foresee, so that a blind execution of an order can also lead to consequences which they never wanted or intended. Therefore, the subordinate also has responsibilities which can never be fully handed over to the superiors. Instructions may come into conflict with other statements given by the supervisors, or they may be contrary to law or moral principles such as the violation of human rights. Blind obedience can also lead to incorrect results when circumstances arise that no one foresaw but which the subordinates ignore for fear of being punished by superiors for deviating from the order.

In the trials of war criminals after the Second World War many of the accused defended themselves by stating that they were "just following orders". Some courts did not follow this reasoning and convicted the defendants of murder or of being accessory to murder because they had a duty to refuse to carry out immoral commands. In a startling judgment, the German Federal Administrative Court in 2005 found that a soldier who refused to participate in acts of war which he deemed illegal was right in his judgment. The court found that soldiers need to check an order to determine whether it is compatible with human rights. The court ruled that not even soldiers are required to submit to absolute authority but only to practice a "reflective obedience" (in German: *mitdenkender Gehorsam*) (verdict of the 2nd Military Service Senate of June 21st, 2005, German Supreme Administrative Court 2 WD 12.04).

Stanley Milgram has shown in his famous obedience experiments that people are ready to commit terrible acts when responsibility is transferred to an authority. He asked his subjects to carry out a learning experiment with another person. Whenever the learners made a mistake, they should give them an electric shock, which at the beginning was only a few volts, but was increased in successive steps until the shocks reached 450 volts, thus apparently inflicting severe pain upon the learner at the end of the test (The participants did not know that the "learner" was an actor who in reality received no shocks.) The authority in these

experiments was only a scientist who told the subjects that they would help science with their cooperation. In the first experiment, over 60 percent of participants went all the way to the end, even though they would have faced no clear disadvantages if they had stopped prematurely. In another experiment the shocks were even higher. Milgram attributed the behavior of the shock-givers to the social circumstances. (Milgram 1974)

Must we therefore accept that such atrocities are caused by the social environment? This would leave us with the sole possibility of changing the circumstances in order to make the whole society "moral" before the individual is moral. But it seems that we do not need to wait for this to occur. The fact that many subjects were not willing to submit to the authority to the bitter end shows that there may be internal factors which can reduce blind obedience. Milgram himself noted that people with higher education and those who deal with people professionally were more likely to refuse obedience than those with lower education or than technicians and engineers. Therefore more and better education could contribute to instilling greater personal responsibility.

Lawrence Kohlberg repeated the Milgram experiments and demonstrated that obedience to authority is highly dependent on the moral competence of the subjects. Among the test subjects with high moral competence many refused to carry out the test to the end, in contrast to the subjects with medium or low moral competence, almost all of whom followed the leader's instructions to the bitter end for the "learner" (Kohlberg 1984). Kohlberg's experiment thus shows that a lack of moral competence can cause people to submit to the judgment of an authority, rather than to judge for themselves. They tend to follow the instructions of powerful institutions rather than their own conscience. On the other hand, people with higher moral competence do not blindly follow orders. All Stage-5 subjects showed civil courage and disobeyed the orders before the maximum voltage was to be administered.

The school is also an institution based on rules and instructions. The legislature provides the legal framework, the state ministry of education or another administrative body formulates the learning objectives and the curriculum and the teacher tells the students what and how they should learn. All of these arrangements are usually well thought out and have been reconsidered and reassessed numerous times. Nevertheless, they can be counterproductive in the event of uncritical and automatic implementation: students learn less than they should, develop learning displeasure or even an aversion to society, although society invests a lot of money in the education system to ensure that children are well prepared for life and that society works well in the future. That is to say that we should also expect *critical* reasoning and reflective obedience from teachers and students. They must – like all other people within the school system – take on a *shared responsibility*. Unfortunately, far too rarely do teachers and students make full use of their existing rights to participation. (Lind & Raschert 1987)

Reflective obedience is an even a bigger challenge to the moral competence of soldiers. On the one hand the lives of soldiers or the fate of a nation can depend on everyone's obedience to commands. But if these commands violate human rights the soldiers must have the freedom to disagree. How else can soldiers be motivated to commit themselves to the defense of freedom and democracy if they cannot practice these principles in their immediate social environment? This consideration has convinced the leadership of the German Armed Forces (*Bundeswehr*) to promote the moral competence of soldiers through the KMDD (Bergmann 2007). The then Inspector General of the Armed Forces commissioned me to train officers, chaplains, and military psychologists with the KMDD method.

The officers delegated to participate in the training appeared at first to be surprised, and some were also irritated. "There is no place for democracy and debate in the military. We only deal with orders and obedience," an officer said, to the applause of his comrades. In response, I

could have cited the "Joint Service Regulation 10/1, Section 301" [*Innere Führung*], which was known to all the participants, but apparently not present in their minds: "Through the *Innere Führung* (inner leadership) the values and norms of the Constitution (*Grundgesetz*) are actualized in the German Armed Forces. It represents the principles of freedom, democracy, and the rule of law in the armed forces. Your mission statement is "citizens in uniform" (Federal Ministry of Defense 2010). Instead, I told them the fictitious story of Corporal Snyder who was on the night shift, guarding an ammunition depot in Afghanistan (I owe this story to Roland Wakenhut 1982). Snyder sees a figure running along the perimeter of the enclosure fence. Despite two warnings, the person does not stop. He wonders, "Is it some enemy stealing ammunition or only a drunken comrade? It's too dark to shoot them cleanly in the legs. My shot could be fatal! What shall I do?"

When I had finished presenting the story, the officers immediately started a lively discussion! Apparently there is a need for thought and discussion even among soldiers. From that moment on, we had a very intense weeklong workshop. At the end, participants informed me that they could not wait to perform KMDD sessions with their draftees and company officers. There was just one problem, however: the garrison commandant could come into the class at any point. As an officer put it, "If just one participant said something that did not suit the commander, I could lose my stripes right away." I advised him to invite his commander to be his supervisor. In other cases this has solved the problem.


Academic learning and teaching

The above mentioned experiments already suggest that moral competence would also have an impact on academic and other learning. In fact, Horst Heidbrink showed in a learning experiment that students with high MCT C-scores can grasp more information from an instructional video than students with low C-scores (Heidbrink 2010). But at first we did not know how to make sense of this finding. Eventually we realized that it fits in well with our growing knowledge about the nature and significance of moral competence. It confirms Piaget's theoretical position that affective processes are involved in all learning, including mathematics, and moreover, that affective learning precedes cognitive learning (Piaget 1976; 1981). It also confirms Daniel Goleman's (2000) theory of 'emotional intelligence'.

I have found additional empirical confirmation for the affective-cognitive connection, including a small study with teachers who participated in a further education program. It turned out that all student-oriented classroom methods – such as the project method, cooperative learning method, team-teaching, and others – were positively assessed by the teachers, but few of these methods were actually applied in their own teaching. These few also demonstrated a high moral competence in the MCT. Apparently, teaching methods in which students can get involved with their own ideas and interests, place a high demand on the moral competence of teachers. If more than one person can determine how and what is learned at school, there are inevitable conflicts, which only do not represent a problem if the teacher has learned to deal with such circumstances. This means that moral competence also plays an important role in academic and professional learning. This finding suggests that if we want teachers to use student-oriented methods of instruction they should be given an opportunity to develop their moral competence.

Conversely, a lack of moral competence can hinder learning. If students carry moral conflicts that they cannot solve around with them then this is likely to greatly retard their learning. The solution of a moral problem is felt more urgently than the solution of academic tasks that should be worked on for school. Often little space is left for academic learning when the mind

works intensively on addressing moral problems. Conversely, this means that one can also influence academic learning by fostering moral competence among students. In fact, various unpublished studies show positive relationships between morality and test scores in almost all subjects (see below, Chapter 6).

This is also true for teachers. In a small study of fifteen experienced teachers, I asked them what they think about modern student-oriented instruction methods like cooperative learning and student projects methods. They all welcomed these and other student-oriented methods. But when I asked how often they actually use cooperative learning methods in their classes, most said "never." A few teachers said "sometimes", "often" or "very often". More, but not all, said they would use student projects as a teaching method. In both cases, actual use of students-oriented teaching correlated strongly with the teachers' moral competence score:
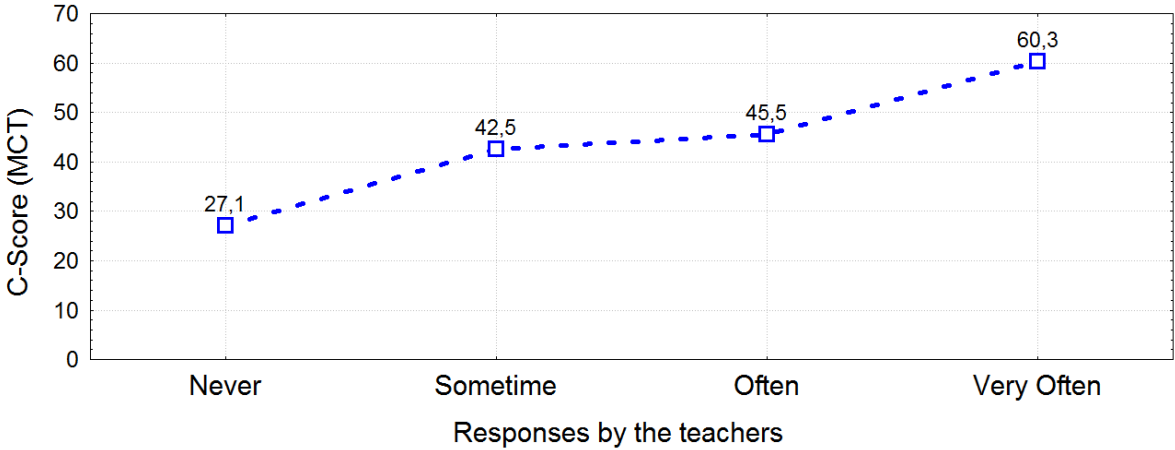
Moral competence and teaching style: how often do you use *cooperative learning* in your class? These findings stem from an unpublished survey of 15 German high school ("Gymnasium") teachers.
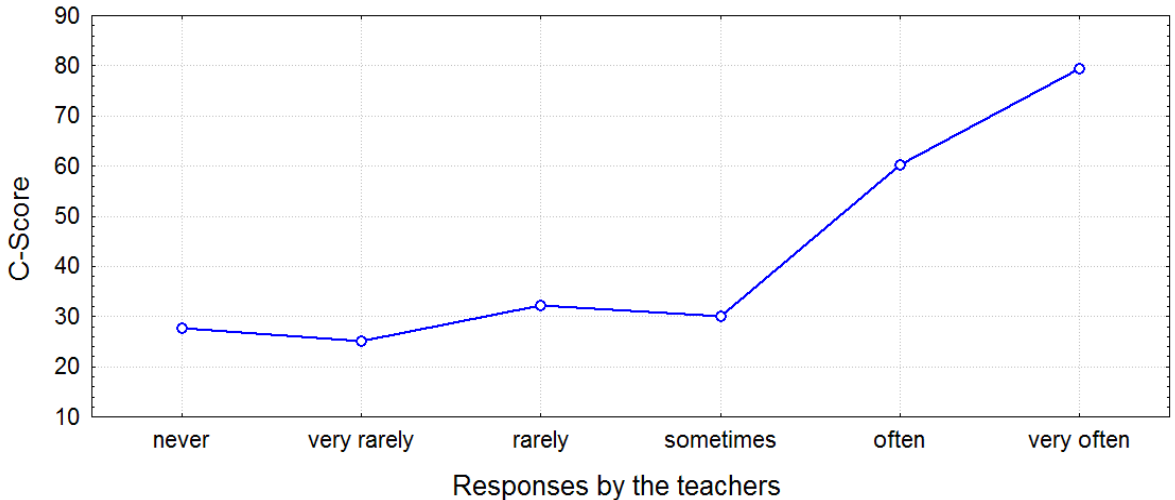
Moral competence and teaching style: how often do you use *student projects* as a method in your class? Source: see figure above.

We should be cautious about drawing too strong inferences from this small survey. However, it is astonishing how high and consistent the correlations were, also in regard to other teaching methods (discussion sessions, peer teaching, Montessori-type free learning). The findings suggest that all teachers would like to involve students more in decisions about the methods, content and aim of learning. But many seem to feel overwhelmed by the expectation that such

involvement would also mean more conflicts in the classroom, which they could not handle. This means that we could improve classroom teaching if we promoted the moral competence of teachers instead of just urging them to use better teaching methods.

Reducing pluralistic ignorance and social stereotypes

Pluralistic ignorance and social stereotypes seem to be directly linked to moral competence. They seem to be caused by a lack of communication about things that are difficult to talk about, or a lack of opportunities to get to know other people. If you, dear readers, believe I think something which I actually do not think, this could be a case of what social psychologists call "pluralistic ignorance". David Krech and Richard Crutchfield describe this phenomenon as follows "No one believes, but everyone believes that everyone else believes" (Krech & Crutchfield 1948, p. 389). It has to do with whether or not people's opinions of other individuals are perceived relatively accurately or whether they are rather based on stereotypes that have little to do with reality. Ignorance and stereotyping usually represent a major burden for social relationships. All too frequently stereotypes are a direct cause of social conflict and violent confrontations and of racism and xenophobia. Stereotypes about other people apparently lower the threshold for starting a war or for cheating others. It was not by chance that the research on stereotypes and pluralistic ignorance had its roots in the United States at the time of racial segregation in the 1920s. Investigations revealed that stereotypes about 'colored' citizens played an important role in the refusal of the white Americans to allow their African American fellow country people full civil rights. Research also shows that mutual stereotypes are the greater the less people have contact with each other. This creates a vicious circle: the separation of people in residential areas, public parks, buses and schools reinforces the stereotypes, which in turn cause further segregation into separate areas. The social psychologists Carolyn and Muzafer Sherif showed just how spatial separation can trigger stereotypes and aggression through simple experiments in camps for youth. Arbitrarily dividing the young people into two groups by drawing a line between them on the floor was already enough to trigger hostile feelings toward the other group. (Sherif & Sherif 1969)

How stereotypes and pluralistic ignorance impact behavior can be seen, for example, when studying helping behavior. Whoever thinks that their fellow human being is not helpful is even less willing to offer help to others. Someone who believes that an acquaintance has deliberately not greeted them may no longer greet another the next time. Or we can look at the school. In one study, teachers told that they believe students learn reluctantly, although when they themselves are asked students place great importance on learning. Teachers associate students' desire for more democratic participation with their alleged dislike of rules. In contrast, students associate the desire for more democratic participation with greater appreciation for rules. (Lind 1995; 2002) One can easily see why students feel misunderstood by their teachers and find it hard to build a relationship of trust with them.

For a long time it was assumed that stereotypes are natural and therefore unalterable, and that we must then resign ourselves to a life of misunderstanding and troubled coexistence. At most, it was assumed that it was possible for people to learn not to let their prejudices get through to the outside by threatening them with disadvantages or penalties. There were rules set for correct language ("political correctness"), which should ensure that people are not offended by stereotypical descriptions. But was that enough? Stereotypes can also be effectively concealed in such a way that penalties are rendered ineffective. What if most people in a society have certain stereotypes and yet there is no one who can enforce the 'correct language'?

Two findings of experimental moral psychology give cause for hope that a sustainable reduction of stereotypes and pluralistic ignorance is possible. Wolfgang Wasel revealed in a series of experiments that people can learn not only to suppress their stereotypes after they have surfaced but also to prevent from developing. Social psychologists give this method of learning the somewhat unusual name of "chronification of fairness goals". By this they seem to mean the same thing as we do when we speak of training moral competence. This comparison is close because, as Wasel writes, "this objective [...] first of all requires a lot of effort and [...] shall only apply if sufficient capacity is available." (Wasel 1997, p. 136; see also Moskowitz et al. 1999) That is, the overcoming of negative stereotypes and prejudices requires an intensive learning process.

The second finding, which provides hope, comes from the experience of the Just Community method (JC) in schools where students have more opportunities to discuss serious topics with each other than in normal schools (Chapters 10 and 11). Ann Higgins and her colleagues describe how pluralistic ignorance with respect to helping behavior considerably decreased in JC schools. While at the beginning of the JC project most students stated that they believed others were not helpful (even though most say that they were helpful), after a year it was only a minority who still believed this. (Higgins et al. 1984)

We observed similar changes in the project *Democracy and Education in Schools* (DES; Section 11.4). A case illustrates some of these changes. Some teachers complained about the students in a Just Community event including both teachers and students. They would constantly break the rule according to which students must not leave the school grounds during the recess breaks. "The students do not seem to really care about rules," said a teacher. As this opinion was delivered openly, the students then had an opportunity to comment. They said that the school rules are not meaningless to them at all, but that some students received no breakfast from their parents and they had to buy something to eat. This discussion led in the end to a constructive solution: the assembly decided that the students be allowed to set up a kiosk on the school grounds, where you could buy food and drinks without having to leave the premises. Based on our evaluation data, we know that intensive talks with one another in the course of the two years in which the project ran resulted in a measurable increase in the students' moral competence. There also appeared to be major changes in the teaching staff, as a school administrator reported the development of an "egalitarian communication culture" that was maintained even beyond the end of the DES project (Section 11.4).

A one-way communication structure in classrooms favors pluralist ignorance and prejudices among teachers, which is detrimental to the relationship between teachers and students, and thus also to learning (Oser 1986). Teachers who only talk in class themselves and give students little opportunity to speak up often know little about their students. So they must rely on prejudice and presumptions in assessing their students. KMDD sessions provide teachers time and ease to listen to their students. The remarks made by a Latin teacher about his students show just how well the method assists teachers to reduce pluralistic ignorance and prejudice about their students: "I have noted with astonishment that some students, who have never been involved in my classes, participated very actively in the KMDD session. I thought they were brain dead." I have often heard such remarks by teachers. Obviously they are based more on pluralistic ignorance than on unprejudiced observation of students.

Sizing up the moral competence of others

Individuals with higher levels of moral competence seem better able to size up the moral ability of their counterparts, at least those whom they know well through work and leisure.

Wolfgang Wasel (1993) found confirmation of this thesis in the already mentioned simulation experiment. Although none of the participants could simulate a higher moral competence than they had themselves (as we had predicted on the basis of the competence theory of morality) participants with high moral competence could simulate familiar individuals' moral competence more accurately than the participants who had a low C-score.

Commitment to a democratic way of life

As hinted at in the previous sections, the causality of moral competence and behavior is often bilateral. Not only is moral competence relevant for behavior in many areas, but conversely, the opportunity to think and discuss leads to an increase in moral competence. By being used to solve problems and conflicts, moral competence grows much like a muscle that is needed for sports, thereby becoming stronger and more agile the more it is used. There are indeed indications that moral competence is a necessary condition for working on the preservation and development of democracy and that the reverse is also true, namely that intensive and equal participation in democratic processes can lead to an increase of moral competence.

The role moral competence plays in the preservation of democracy was evident in the emergence of the Free Speech movement in the 1960s in the United States which kicked off a world-wide student movement. This emerged in a climate of suspicion and restriction of freedom of expression. The racial segregation and the Vietnam War also played a major role. But the situation which triggered the movement off was, as is often the case, rather negligible. According to various reports, some students at the University of California at Berkeley decided to protest against the increase in prices in their cafeteria. The university president forbade the demonstration and called the police. The students insisted on the observation of the terms of the First Article of the US Constitution, which guarantees every US citizen freedom of expression. The rector then declared that the Constitution did not apply on campus because it is private. Now even more students demonstrated, but this time for the fundamental right to freedom of speech. Norma Haan and her colleagues interviewed several hundred participants and non-participants with a written version of Kohlberg's *Moral Judgment Interview*. They found that there was significantly more participation of students with high moral competence among the demonstrators than in a comparison group of non-demonstrators (Haan et al. 1968). In a study with German students a clear connection between moral competence and democratic value systems was also found. (Gielen 1986; Lind et al. 2010)

We also found support for the assumption that longer, active participation in discussions in a democratic movement can be beneficial for the development of moral competence among Polish students who participated along with German, Dutch, Austrian and Slovenian students in our international comparative longitudinal study from 1977 until 1985. (Bargel et al. 1982) While the moral competence developed only slightly in the students in the other four countries, my colleagues and I found a strong increase in Polish students in the first two years. Ewa Nowak and I re-analyzed these data (Nowak & Lind 2009). We found evidence that this rapid development was due to the fact that most participants of our survey actively participated in the discussions and protests during that time. It was the time of the Solidarność-Movement, in which many people campaigned for greater democratization of Poland. When the students were asked again after two years, the moral competence of Polish participants had fallen back to the original level. A new government had in the meantime taken over power, declared martial law and banned demonstrations and discussions. We believe that this decline in moral competence was a consequence of the martial law which left no room for public discussions anymore, so that the students' moral competence received no stimulation.