ETHICS IN PROGRESS

12/1

Ethics in Progress

Vol 12, No 1. 2021.
Peer-reviewed online journal.

A multidisciplinary forum for research bridging between academic ethics and social practice.

EDITORIAL TEAM

Vol. 12 (2021) No.1

Table of contents:

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #1, pp. 4-5. DOI:10.14746/eip.2021.1.1

Introduction

Roma Kriaučiūnienė

(Vilnius University; Lithuania; roma.kriauciuniene@flf.vu.lt)
ORCID: 0000-0002-9356-1098

It is a pleasure for me to introduce papers presented at the 14th International Symposium „Moral Competence: Its Nature, Relevance, and Education" hosted by the Institute of Foreign Languages of the Faculty of Philology at Vilnius University, Lithuania, on 23-24 July 2020, in collaboration with several academic institutions from abroad, such as Adam Mickiewicz University in Poznań, Poland. Due to the pandemic restrictions, a symposium was organized in an online mode. Six papers were revised and submitted to this special issue as symposium proceedings, completed by two additional and related papers. The contents offer a thorough insight into the concept of – and reasearch into – the moral competence defined by Georg Lind (Institute for Moral-Democratic Competence; formerly: University of Konstanz) and visualized by his Moral Competence Test (formerly: Moral Judgment Test) certified in 41 languages. In certain papers research findings and methods based on further measuring instruments, as well as the alternative approaches to moral judgment and decision making were applied and discussed (e.g., moral foundations approach).

Roma Kriaučiūnienė's (Vilnius University) and Georg Lind's (University of Konstanz and Institute for Moral-Democratic Competence in Konstanz) report on validation study and certification of the MCT for the Lithuanian language opens the volume as the most recent contribution to the core topic. Lind's paper "Making Human Traits Visible, Objectively *and* Validly, Through Experimentally Designed Questionnaires" provides a transparent 'demarcation line' between classic psychometric tests and the MCT. Papers authored by Bogdan Popoveniuc (University Stefan cel Mare of Suceava/West University of Timişoara), Marina Klimenko (University of Florida, Gainesville), Aida Mofakhami (Allameh Tabataba'I University, Teheran), Georg Lind (University of Konstanz and Institute for Moral-Democratic Competence in Konstanz), Marina Klimenko (University of Florida, Gainesville), Kamila Stastna (Friedrich-Schiller-Gymnasium in Pirna), Ewa Nowak (Adam Mickiewicz University in Poznań), Anna-Maria Barciszewska (Karol Marcinkowski

University of Medical Sciences in Poznań), Kay Hemmerling (Arbeit und Leben Institute in Leipzig), Sunčana Kukolja Taradi (University of Zagreb) depict moral competence development in various contexts (for instance, education and gender in Klimenko, conformity behavior in Mofakhami, and medical education in Nowak, Barciszewska, et al.). Complementary or alternative approaches (e.g., moral foundations in Popoveniuc, Stastna and Alexandra Wasielewska [Adam Mickiewicz University in Poznań) follow extend and sometimes also challenge our research topics.

Acknowledgements:

We wish to express our sincere gratitude to Prof. Ewa Nowak, and Prof. Georg Lind for their sincere and valuable guidance extended to us in the process of the organisation of the 14th International Symposium „Moral Competence: Its Nature, Relevance, and Education". We are also very grateful to our colleagues Agnė Vaidakavičiūtė and Remigijus Ptičkinas for their professional assistance with many technical issues of our work for the Symposium.

We express our deep sense of gratitude to Prof. Ewa Nowak who lent her professional guiding for the publication of this special issue.

We take this opportunity to thank Stephen Dersley, Marcin Byczyński, Małgorzata Steć, Jairo Ordóñez, Michael Gommel, Margret Ruep, Kay Hemmerling, Anna Malitowska and Mateusz Bonecki for diverse organizational support.

Last, but not the least, we express our gratitude to all the authors for their contribution to this publication.

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #2, pp. 6-14. DOI:10.14746/eip.2021.1.2

Validation of the Lithuanian Version of Moral Competence Test and Application of KMDD in Language Teaching / Learning Process at Vilnius University, Lithuania

Roma Kriaučiūniene

(Vilnius University; Lithuania; roma.kriauciuniene@flf.vu.lt)
ORCID: 0000-0002-9356-1098

Georg Lind

(University of Konstanz; Germany; georg.lind@uni-konstanz.de)
ORCID: 0000-0002-2235-5465

Abstract: The findings obtained by G. Lind using his original research instrument – the Moral Competence Test – suggest that universities lack the capacity to foster students' moral competence development. The MCT has been translated into 39 languages, all of which have gone through the necessary validation procedure. The article reports on the MCT validation study for the 40th language, namely Lithuanian. The research sample consisted of 526 students of English/German/French languages, future foreign language teachers, in the 1 st to 4th years of study at two universities in Lithuania: the former Vilnius Pedagogical University and the Vilnius University. The majority of the respondents demonstrated low or medium level of moral competence. On the basis of this cross-sectional study (2019–2020), the MCT for Lithuanian has been successfully validated and certified. In the following article, we present and discuss all the validation criteria and revisit the theoretical background of MCT. We also argue for educating students in moral competence and evaluating the effects of moral competence promotion in academic contexts.

Keywords: Moral competence; moral competence development; Moral Competence Test in Lithuanian version; MCT; validation study; KMDD; Konstanz Method of Dilemma Discussion; English language teaching/learning process.

1. Objectives

The following paper reports on the validation and certification of the *Moral Competence Test* (MCT) for the Lithuanian language. Moral competence is defined by Lind (2019) as the ability to resolve problems and conflicts on the basis of moral principles, merely by thinking and discussion, instead of through the use of violence, deceit or by submitting to an authority. In particular, it is defined as the ability to evaluate arguments put forward by others in terms of their moral quality rather than in terms of opinion agreement. This is the very prerequisite for living in a democratic society.

The MCT is a behavioral experiment which makes moral competence visible. It is

based on the psychological dual-aspect-theory of moral behavior, but does not invoke any statistical assumptions as other psychological tests do. In order answer the question of whether the test measures what it is supposed to measure, the MCT was submitted (a) to several reviews by experts of moral psychology, and (b) to three rigorous empirical tests. These three tests were also used to check the cross-cultural validity of the Lithuanian version of the MCT. In addition, we report the findings from a study on the significance of moral competence for a teacher's ability to foster students' moral competence.

The validation study was conducted with students (N=526) of Vilnius University in 2020. The Lithuanian MCT conforms almost perfectly with the first two criteria. Because of this nearly perfect confirmation, the correlations were so high that the third criterion could not be tested.

Moreover, the findings indicate the distribution of moral competence in Lithuanian students training to be language teachers, in terms of their development during their university courses.

2. Theoretical Background

Lind (1978; 1982; 2019) created the MCT in order to make moral competence visible. It is a behavior experiment with a multivariate orthogonal design, as Egon Brunswik (1955) had suggested. This means that the MCT is not a "test" in the sense of psychology tests, but it is an *n=1 experiment*, operationalized as a questionnaire. Because of this experimental design, the participants' pattern of responses lets us directly *see* the properties of their moral competence without the aid of certain statistical models, as is the case with classical psychological tests. As a result, the criteria of classical tests do not apply. The moral consistency of participants' responses is used as an indicator for their moral competence, not as a sign of test stability. As the MCT has not been changed since its inception in 1977 (Lind 1978, 2013), it can be considered stable and reliable.

The MCT's validity is checked using four well established psychological findings about the nature of moral judgment behavior:

1) The six types of moral orientations, as defined by Kohlberg (1984), form a universal preference hierarchy;

2) As Piaget (1976) concluded from his observations, moral orientations and moral competence are "parallel", that is, they correlate in highly predictable manner;

3) Neighboring types of moral orientations correlate higher with one another than more "distant" types (simplex structure); and

4) Moral competence cannot be faked upward, as attitudes and orientations can. These psychological (instead of solely statistical-formal) validity criteria provide much more rigorous criteria for the validity of experimental designs than the conventional statistical criteria used in test psychology. These findings serve as criteria of the MCT's psychological validity. The first three are obligatory for testing the validity of new test

versions and translations of the MCT.

These three criteria are very rigorous. They allow us to detect even small errors. Rigorous means that the *a priori* probability of confirming these criteria by chance is extremely small, and, therefore, their confirmation is extremely informative (Karl Popper). For example, the *a priori* probability of confirming the statistical hypotheses that two groups of people differ regarding some trait, is almost 100%, if the groups are very large. In contrast, the *a priori* probability for a single individual's preference hierarchy of six moral orientations is only 1 / 6! = 1 / 720 = 0.0014 or 0.14 %. When we test this validity hypothesis with many individuals, the probability of a confirmation by chance is extremely small.

Although these validity criteria are very rigorous, they have been supported very well so far by all studies. Hence, we can safely say that the MCT *is psychologically valid*, i.e., it can truly make moral competence visible.

3. Method

The Moral Competence Test (standard version) consists of two dilemma stories. Each story is followed by six arguments in favor of the protagonist's decision, and six against it. Each argument represents of the six types of moral orientation defined by Kohlberg (1984). The first author of this paper has translated the MCT from English into the Lithuanian language. Participants were recruited and surveyed randomly. Participation was voluntary and anonymous, occurring during their class hours. Data were collected for several weeks. The sample consisted of 654 participants, 604 females and 50 males, aged 19 to 23, only Lithuanian speakers, representing various fields of study. The central C-score was computed according to the algorithm developed by Lind.

Here is one of the dilemma stories and some sample arguments:

MORALINĖS KOMPETENCIJOS TESTAS

Prašome atidžiai perskaityti šias dvi dilemas. Įvertinkite pasiūlytus jų sprendimo variantus ir prie kiekvieno iš jų pažymėkite, ar Jūs sutinkate ar nesutinkate su tokiu sprendimu apibraukdami skaičių nuo -4 iki +4.

Darbininkų

Dėl atrodytų nepagrįstų atleidimų iš darbo kai kurie gamyklos darbininkai įtaria, kad vadovai, naudodamiesi vidinio telefono ryšiu, slapta klausosi savo darbuotojų pokalbių ir šią informaciją naudoja prieš juos. Vadovai oficialiai ir primygtinai neigia šiuos kaltinimus. Profesinė sąjunga teigia besiimsianti priemonių prieš kompaniją tik tuo atveju, kai šie įtarinėjimai bus pagrįsti tikrais įkalčiais. Tuomet darbininkai įsilaužia į administracijos patalpas ir paima įrašus, įrodančius, jog kaltinimai dėl slapto klausymosi yra pagrįsti.	Ar Jūs pritartumėte ar nepritartumėte darbininkų elgesiui?						
	gana teisus						visiškai neteisinga
	-3	-2	-1	0	+1	+2	+3

Ar pritartumėte šiems argumentams *palankiai* vertinantiems dviejų darbininkų elgesį? Tarkime, kas nors tvirtintų, jog jie buvo teisūs...	Visiškai nepritariu								Visiškai pritariu
1. nes jie nepadarė jokios žalos gamyklai.	-4	-3	-2	-1	0	+1	+2	+3	+4
2. nes įmonė nepaisė įstatymų, vadinasi, dviejų darbininkų elgesys buvo leistinas teisingumui ir tvarkai atstatyti.	-4	-3	-2	-1	0	+1	+2	+3	+4
3. nes dauguma darbininkų pritartų jų poelgiui ir dauguma tuo džiaugtųsi.	-4	-3	-2	-1	0	+1	+2	+3	+4
4. nes pasitikėjimas žmonėmis ir asmeninis orumas reiškia daugiau negu įmonės gerovė.	-4	-3	-2	-1	0	+1	+2	+3	+4
5. nes įmonė pati pirma pažeidė teisingumą, tai dviejų darbuotojų įsibrovimą į administraciją galima pateisinti.	-4	-3	-2	-1	0	+1	+2	+3	+4
6. nes abu darbininkai nematė jokių juridinių priemonių, kaip atskleisti įmonės piktnaudžiavimą ir todėl pasirinko, kas jų manymu teisinga.	-4	-3	-2	-1	0	+1	+2	+3	+4
Ar pritartumėte šiems argumentams *nepalankiai* vertinantiems dviejų darbininkų elgesį? Tarkime, kas nors tvirtintų, jog jie buvo neteisūs...	Visiškai nepritariu								Visiškai pritariu
7. nes įstatymams ir tvarkai visuomenėje iškiltų grėsmė, jeigu visi elgtųsi taip, kaip tie du darbininkai.	-4	-3	-2	-1	0	+1	+2	+3	+4

(Excerpt of the Moral Competence Test in Lithuanian translation; the full test is available from the second author: Georg.Lind@moralcompetence.net).

4. Results Regarding the Validity of the Lithuanian MCT

The findings confirm the first two hypothesis very well.

The criterion of "preference hierarchy". The six types of moral orientations are preferred and rejected in the way that moral psychological theory predicts: The moral orientation types 5 and 6 are the most preferred, and and types 2 and 1 are the most rejected (Figure 1). There is only a slight inversion of preference order among the orientations types 1 and 2. So far, this has been found in all other studies (Lind 2002, 2013).

The criterion of "affective-cognitive parallelism". The six types of moral orientations correlate with moral competence (the C-score of the MCT) exactly in the way that was predicted on the basis of Piaget's notion: The type 6 moral orientation correlates the highest with moral competence, and type 1 the lowest. In other words, the higher the participants' moral competence, the more clearly they prefer better moral orientations than less adequate ones.

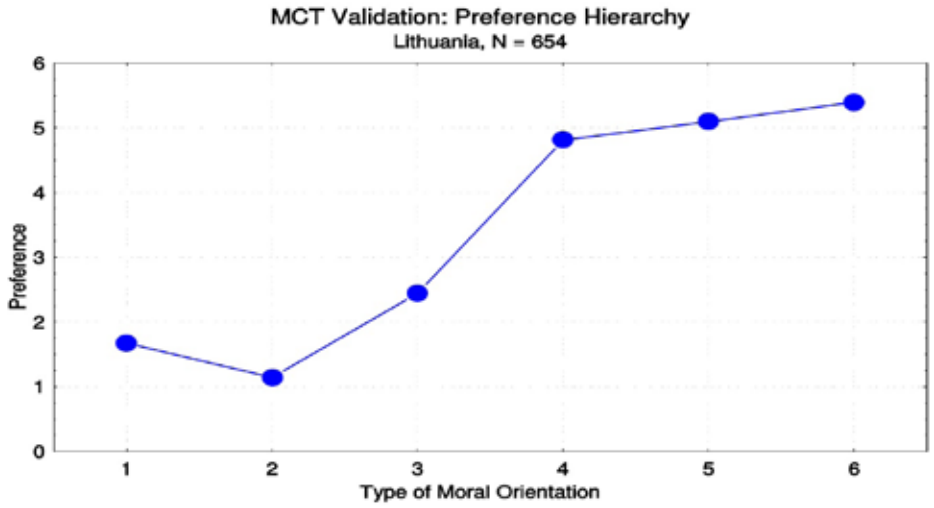


Figure 1: Validation criterion 1: Preference hierarchy (moral orientations 1 to 6, according to Kohlberg). This criterion was fully met.

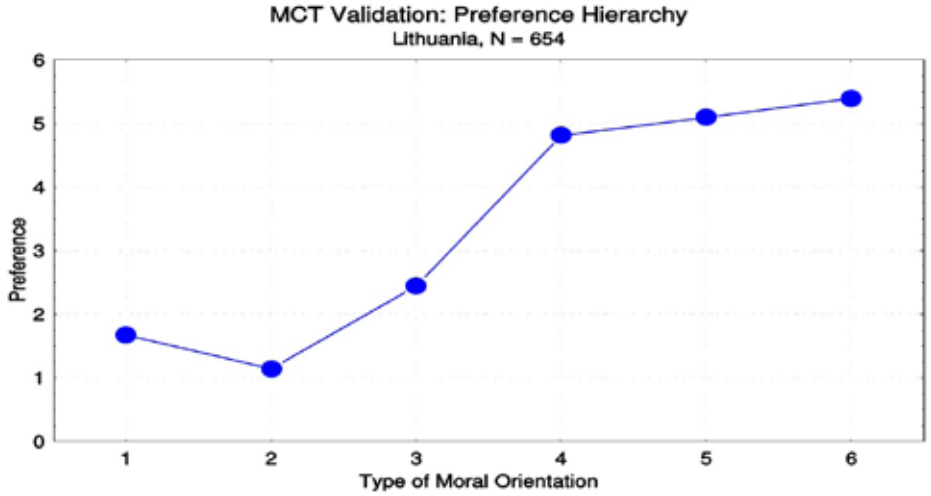


Figure 2: Validation criterion 2: Affective-cognitive parallelism. This criterion was fully met.

The third validity hypothesis (simplex quasi-structure) could not be tested in this data set because the first two critera fit extraordinarly well. Because of this, all six types of moral orientation correlated almost perfectly with one-another. This does not imply that the third criterion of validity is falsified, but only means that it cannot be tested with this set of data. The fourth criterion is optional, because it is very laborious.

In sum, like the original MCT, the Lithuanian translation is highly valid. That is, it allows us to measure two distinct aspects of moral behavior, moral orientations and moral competence, simultaneously and validly (1982, 2002, 2019). Hence there is no need anymore to rely on subjective methods like clinical interviews (Lind 1989). The MCT makes moral competence visible without imposing statistical models on the data. Statistics are used only to translate the graphical results into numerical results (C-score) so that the findings can be used for studying research questions.

5. The Relevance of Moral Competence for a Teachers' Ability to Foster Moral Competence

Many research findings suggest that universities fail to foster students' moral competence (Lind 2000, 2001, 2002, 2013). Therefore, we tested the moral competence of 526 students of English, German and French languages, future foreign language teachers, in the 1st – 4th years of their studies at two universities in Lithuania (the former Vilnius Pedagogical University and Vilnius University). All students showed, as seen in Figure 1 above, high moral orientations. They preferred principled moral reasoning (Type 5 and 6) over legalistic or conformity morality, and these over pre-conventional moral orientations.

However, their average moral competence was rather low. Many lack the ability to apply their moral orientations to their behavior.

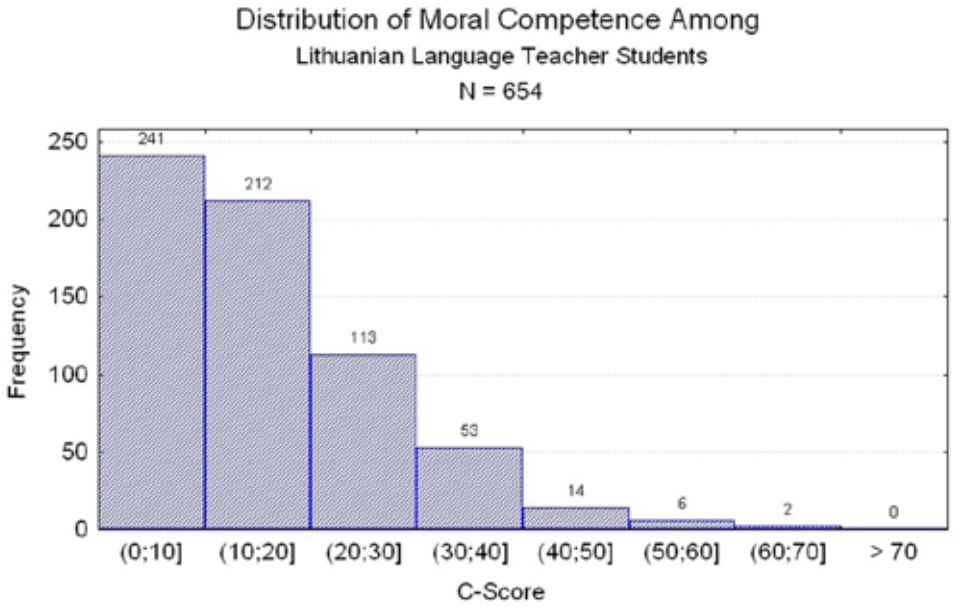


Figure 3: Distribution of moral competence.

Some researchers (Wood 2007; Thornberg 2008; Brimi 2009; Hoekema 2011**;** Kiss & Euben 2011) argue that fostering students' moral development should be given a high priority by universities. In spite of this, universities seem to have little impact on students' moral competence development. As Figure 5 shows, students' moral competence does not seem to be affected at all by their university studies.

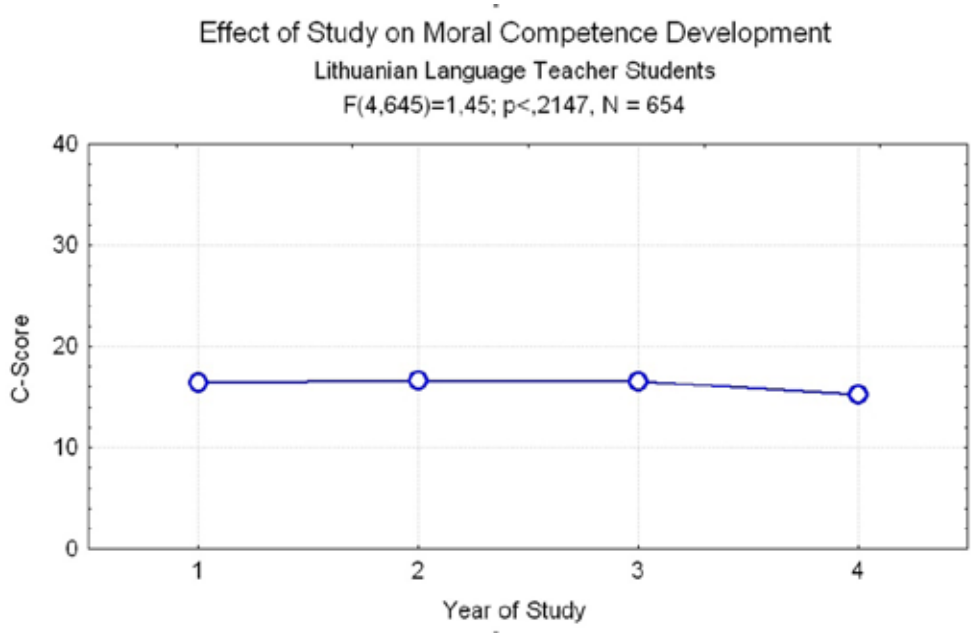


Figure 4: Changes in university students' moral competence during their studies (cross-sectional study).

How could that be changed, ask E. Kiss and J. P. Euben (2011)? Moral competence research suggests that this can be done in two ways. First, students' moral competence can improve when their university gives them at least some opportunities for responsibility-taking and guided reflection (Lind 2000, 2002, 2013, Schillinger 2006). Second, students' moral competence can be promoted purposefully by using the *Konstanz Method of Dilemma-Discussion* (KMDD; Lind 2019). We applied the KMDD in the teaching and learning process of the English language at Vilnius University and the KMDD effect upon the learners' moral competence C-score following the pre-testing and post-testing procedure by MCT was measured. The intervention was carried out in 2019 during the English language course *English for* Academic *Purposes and Research* with 1st year students of the study program *Childhood Pedagogy* (30) at the beginning of the semester. The participants were tested before and after the interventions using the MCT. The intervention produced an increase of 1.8 C-points. This is small, because the KMDD-teacher was not fully trained and certified. But it is encouraging as the control group's C-score did not move at all.

Conclusion

The Lithuanian version of the *Moral Competence Test* (MCT) is highly valid. It is the 40th language version of the MCT. Most have also been validated. The validation of some versions is still pending (see http://moralcomptence.net).

The MCT makes moral competence visible. Its scores show how effective or ineffective our schools and universities are in terms of fostering students' moral competence.

Teachers of secondary and higher education can use it to see how effective they are at fostering the moral competence of their students, and whether they need additional training to become more effective.

References:

Brimi H. 2009. "Academic Instructors or Moral Guides? Moral Education in America and the Teacher's Dilemma," *Clearing House* 82(3):125–130. Academic Search Complete, EBSCO*host*, viewed 5 January 2015.

Hoekema D. 2011. "Is There an Ethicist in the House? How Can We Tell?," in E. Kiss & J. P. Euben (Eds.), *Debating Moral Education: Rethinking the Role of the Modern University*. Durham: Duke University Press.

Kiss E., & Euben J. P. (Eds.) 2011. *Debating Moral Education: Rethinking the Role of the Modern University.* Durham: Duke University Press. Published online: 12 October 2011. Springer Science+Business Media B.V.

Kohlberg L. 1984. *The Psychology of Moral Development. Vol. 2: Essays on Moral Development*. San Francisco: Harper & Row.

Lind G. 1978. „Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts," in G. Portele (Ed.), *Sozialisation und Moral* (pp. 171–201). Weinheim: Beltz,

Lind G. 1982. "Experimental Questionnaires: A New Approach to Personality Research," in A. Kossakowski & K. Obuchowski (Eds.), *Progress in Psychology of Personality* (pp. 132–144). Amsterdam: North-Holland.

Lind G. 1989. "Measuring Moral Judgment: A Review of *The Measurement of Moral Judgment* by Anne Colby and Lawrence Kohlberg," *Human Development* 32:388–397.

Lind G. 2000. "The Importance of Role-taking Opportunities for Self-sustaining Moral Development," *Journal of Research in Education* 10(1):9–15.

Lind G. 2001. "From Practice to Theory. Turning Teacher Education on Its Feet." http://moralcompetence.net/b-liste.htm

Lind G. 2002. *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung*. 2nd Edition. Berlin: Logos Verlag.

Lind G. 2013. "Favorable Learning Environments for Moral Competence Development – A Multiple Intervention Study with Nearly 3.000 Students in a Higher Education Context," *International Journal of University Teaching and Faculty Development* 4(4):173–192.

Lind G. 2010. "Growth and Regression in Cognitive-moral Development in Young University Students," in C. G. Harding (Ed.), *Moral Dilemmas and Ethical Reasoning* (pp. 99–114). New York: Routledge.

Lind G. 2019. *How to Teach Morality. New: Discussion Theater.* Berlin: Logos Publishers.

Piaget J. 1976. "The Affective Unconscious and the Cognitive Unconscious," in B. Inhelder & H. H. Chipman (Eds.), *Piaget and His School* (pp. 63–71). New York: Springer.

Schillinger M. 2006. *Learning Environments and Moral Development: How University Education Fosters Moral Judgment Competence in Brazil and two German-speaking Countries*. Aachen: Shaker Verlag.

Thornberg R. 2008. "The Lack of Professional Knowledge in Values Education, Teaching and Teacher Education," *International Journal of Research and Studies* 24(7):1791–1798. URL: http://dx.doi.org/10.1016/j.tate.2008.04.004. Copyright: Elsevier Science B.V., Amsterdam, http://www.elsevier.com (last viewed on 20 Dec 2020).

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #3, pp. 15-38. DOI:10.14746/eip.2021.1.3

Making Human Traits Visible, Objectively and Validly, through Experimentally Designed Questionnaires

Georg Lind

(University of Konstanz; Germany; georg.lind@uni-konstanz.de)
ORCID: 0000-0002-2235-5465

Abstract: Researchers who need valid and objective data for evaluating their theories or the efficacy of educational methods and programs have to choose between two equally undesirable alternatives: either they can use "objective" methods which have a questionable validity, or they can turn to "subjective" assessment methods with better validity. In other words, while subjective approaches to the study of human traits may be, or really are, valid, they lack objectivity, that is, they may be biased toward the researcher's theory. On the other hand, objective approaches to the study of psychological traits often lack psychological underpinning but are solely designed to fit a certain statistical model. Thus, we cannot know what these instruments really measure.

Here I present a new approach to the study of human traits, which claims to be objective as well as psychologically valid, namely the concept Experimental Questionnaire (EQ). An EQ lets us make traits visible without relying on dubious statistical assumptions. Thus, it makes it possible for the researcher to test the psychological theory underlying its designs. The EQ methodology is not only an idea, but it has been applied for constructing the Moral Competence Test (MCT) and for testing the assumptions about the nature of moral competence which were used to design it. So far, all the studies have clearly confirmed their validity. This makes the MCT suitable for testing hypotheses regarding the relevance and teachability of moral competence, and, therefore, also for evaluating the efficacy and efficiency of educational methods of fostering this competence.

Experimentally designed questionnaires can also be used in other fields of educational and psychological research in which testable theories about the nature of its objects have been developed.

Keywords: Psychological measurement; standardized tests; theory; objectivity; validity.

Prologue

In 1979, I had a chance to meet Paul Meehl, the co-author of the famous *Minnesota Multiphasic Personality Inventory* (MMPI). I admired him for his writings on methodological issues in psychology, like "When to use your head instead of a formula" (Meehl 1957). At that time, I was not sure whether my new idea about psychological measurement

would hold water. After listening patiently to my critique of mainstream psychological testing and my idea that should replace it, he cautioned me. He said that psychologists and educators would hardly welcome this new idea because if they did, they would have to give up the tests with which they make their living. Today, 40 years later, I know he was right. Fortunately he had added "Go on!" That encouraged me to write this paper.

The Persisting Dilemma of Psychological and Educational Measurement

Millions of dollars are spent every year on tests of character, academic abilities, vocational skills, mental disorders and so on, in the hope that their findings help to improve therapy, education and the politics of mental health and education (Gregory 2018, 22). But anyone who seeks the service of psychology (which translates to "science of the mind") faces a persistent dilemma. One has to choose between two opposite approaches to the observation and measurement of psychological traits, both of which have their drawbacks:

"Subjective" (also called "qualitative") psychologists argue that the focus on studying the internal structure of the human mind will indeed provide badly needed insights on the human condition. The human mind, they insist, can be studied only by using *subjective* methods like clinical interview.

In contrast, "objective" ("quantitative") psychologists argue that if psychology wants to be recognized as a science, it must use only *objective* methods of measurement. Notably, both agree that the *internal* factors of the human mind and its *structure* are out of reach for objective measurement. Can psychology really become a science if it spares the direct, objective measurement of its very objects?

For many years eminent scholars have argued that this deficit has prevented psychological and educational research from developing into a real science (Travers 1951; Loevinger 1957; Miller 1962) and from playing a more constructive role in evaluating and improving education (Schoenfeld 1999; Ravitch 2013).

For centuries, psychology and education were part of philosophy and, therefore, the domain of subjective science. Philosophers who focused on the nature of the human mind mostly used subjective methods for studying it. Their methods tended to be *ideographic* (acknowledging the individuality of the person) and *holistic* (taking the whole structure of the individual personality into account).

This philosophical approach to the study of the human mind was challenged in the 19th century by objective psychologists who were at home in physics, biology and medicine. They argued that psychological research must be *nomothetic* (searching for general laws) and *objective,* studying people's behaviors instead of the structure of their mind: "The behaviorist recognizes no such things as mental traits, dispositions or tendencies," postulated Watson (1970/1924, 98), the founder of psychological *behaviorism,* which is

still very influential. He and his followers believe that psychological measurement should focus on behavior instead of on psychological traits: "A test is a standardized procedure for sampling behavior and describing it with categories or scores" (Gregory 2018, 23). Their object is not genuinely psychological but only somehow "related" to psychology: "We define psychological assessment as the gathering and integration of psychology-related data" (Cohen & Swerdlik 2018, 2).

This antagonism of the two approaches has caused a deep "crisis of psychology" (Bühler 1927) which hampers the progress of psychology as a science to this day. As the philosopher Wittgenstein (1953) noted: "In psychology there are experimental methods and conceptual confusion. The existence of experimental methods makes us think we have the means of solving the problems which trouble us; though problem and methods pass one another by." Eminent psychologists agree. Similarly, Graumann wrote: "Theoretical frameworks and methodological convictions are still too divergent, if not partially incommensurable" (Grauman 1960, 146, my transl.). Block (1977) also asserted that "perhaps 90% of the studies are methodologically inadequate, without conceptual implication, and even foolish" (Block 1977, 39). The educational researcher Travers observed "that the rather meager advances made in many areas of psychological measurement during the last 20 years are mainly a consequence of the fact that these areas are staffed mainly by technicians interested in producing useful instruments and not by scientists interested in expanding knowledge" (Travers 1951, 137). The statistician and psychologist Kempf wrote: "What usually is called psychological test theory is actually a statistical theory of item selection in order to produce a test with some desirable features" (Kempf 1981, 3, my transl.). Ten years later, Alan Schoenfeld (1999), former president of the *American Educational Research Association* (AERA) and an accomplished educational researcher and mathematician, complained that still "virtually none of the current assessments in wide use are grounded in well-developed theories of competence" (Schoenfeld 1999, 12). Therefore, he called for a moratorium on standardized testing until this basic issue has been solved. More and more educational researchers, teachers, parents and educational policy makers question the meaningfulness and validity of standardized testing (Amrein & Berliner 2002; Ravitch 2013; Sjoberg 2017; Koretz 2017).

Yet not much has changed. Textbooks on psychological tests and measurement do not respond to any of these complaints (Gregory 2018; Cohen & Swerdlik 2018) In psychological measurement, it seems, we have to choose between Scylla and Charybdis, that is, between a psychological object which cannot be measured through objective methods, on the one hand, and an objective method which rejects psychological objects, on the other.

How can we overcome this impasse? Is it really not possible to study the human mind objectively without giving up its genuine object?

The Critical Role of Theory in Measurement

In everyday life we measure all kind of things by reading a scale like a meter stick without much thinking. But we should remember that before we used the meter we had other means of measurement like our hands, feet or elbows. Usually, we do not give it any thought that measurement is something artificial, that is, something which is based on conventions and theoretical assumptions. But "there is no measurement without a theory and no operation which can be satisfactorily described in non-theoretical terms" (Popper 1968, 62). The theoretical assumptions concern, for example, the stability of the material of which the meter stick is made. A rubber band would not be suitable. If the stick is made out of metal, the surrounding temperature might cause the stick to constrict or expand and thus bias our measurement. Using a thermometer requires that the expansion of the fluid inside the instruments expands strictly proportionally to the surrounding temperature. Research shows that this is true only within a certain range of temperature. Outside this range, the thermometer gives incorrect numbers.

The same is true for psychological measurement. When we "read" people's intelligence, morality, political attitudes etc. from their visible behavior, this reading is also based on theoretical assumptions, namely assumptions about the relationship of observable behavior to the things we are interested in. As I will discuss below, even when the relationship between a certain behavior and a certain trait looks simple as in the case of classical attitudes scales, attitude tests produce ambiguous data. For example, a score in the middle range of a conservatism scale can mean that the participant has a "middle attitude toward conservatism." But it can also mean that he or she has no attitude toward conservatism at all, or that he or she has a high differentiated attitude (Scott 1968). The relationship between overt behavior and underlying traits can be even more complex when we look at the relationship between answers to a clinical interview and, let's say, a participant's stage of moral development (Lind 1989).

Measurement theories are the link between reality and our theories of reality. If measurement does not provide us with valid data about reality, our thinking and our decisions will be misled by false images of the world. Hence it is essential that measurement theories are testable and that we actually do test them. Only if our measurement provides valid data, can we trust them and use them for examining the empirical truth of hypotheses about the relevance, determinants and teachability of psychological traits.

will now discuss the two main approaches to measurement in current psychology: the "subjective" and the "objective" approaches. As I mentioned, both have severe shortcomings.

The "Subjective" Approach

Subjective psychologists base their measurement on the assumption that our

behavior is determined mainly by unconscious affects and cognitions. In other words, they believe that only through the study of the unconscious level of our mind can we really understand human behavior and make education, therapy, and politics more effective. They also believe that unconscious affects and cognitions cannot be assessed directly but only indirectly, namely through interpreting people's visible performances in certain situations or their answers to the psychologist's questions. Interpretation means that measurement requires researchers to clearly define their object clearly and concisely, and to make assumptions about its nature in behavioral terms, so that these assumptions can be objectively tested. These assumptions or hypotheses should be based on coherent theories which have been tested by different researchers.

A scientific definition of a psychological object should allow us to examine the truth of the assertion that a test is a valid measure of that object. Unfortunately, in psychology the object of measurement is rarely defined in a clear and concise way. Rather definitions are tautological, fuzzy, or evasive definitions and, therefore, do not allow us to judge a test's validity. For example, "intelligence" is often tautologically defined as "what is measured by intelligence tests" (Bailey 1994, 57); or its definition is vague and ambiguous, like this: "One of Sternberg's very succinct definitions of intelligence states: Intelligent behavior involves adapting to your environment, changing your environment, or selecting a better environment"[1]. A psychological definition is not "succinct" if it states only what is "involved." Moreover, if several "definitions" (plural!) are available, confusion is inevitable. A definition is evasive if it encompasses everything a person does: "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (David Wechsler). No test can measure such a "global capacity", but can only assess a small section of it. The claim that such a section is representative for the whole is not testable.

However, there are a few exceptions. Take for example Lawrence Kohlberg's research on the nature of moral judgment competence. He defines it "as the capacity to make decisions and judgments which are moral (i.e., based on *internal* principles) and to act in accordance with such judgments" (Kohlberg 1964, 425; emphasis added). "Morality (...) defined as 'having moral principles' includes (…) inner conformity to them in difficult situations rather than outer conformity in routine situations" (Kohlberg 1958, 3).

Kohlberg's definition is short and clear: (1) It defines the affective aspect of morality in terms of the individual's *inner* moral principles or orientations, (2) It defines the cognitive aspect of morality as the *structure* of his or her overt moral judgments, and (3) It defines both as *aspects of visible action* or *behavior*. His definition of a psychological trait is distinct from most studies of morality (and other human traits): "The trouble with such [studies] is that they describe the person externally in terms of his impact on and relation to his culture and to the observer's values. They will not tell us how the individual thinks, what values he *actually* holds" (Kohlberg 1958, 82; emphasis added).

1 https://thesecondprinciple.com/optimal-learning/sternbergs-views-intelligence/

But how can we make an "inner" trait visible? When Kohlberg started his research, there were only subjective methods available. Kohlberg followed Piaget's lead and developed his clinical interview method, the *Moral Judgment Interview* (MJI). In this, the interviewer confronts participants with several dilemma stories in which the protagonists have to make a presumably difficult decision: Whatever they decide, they violate a moral principle. The participants should say whether they agree or disagree with the protagonist's decision and why. The interviewer follows up their answers to get a rich picture of their reasons, and also probes into counter-arguments: Which reasons could justify the opposite opinion? The answers of the interviewees are recorded, transcribed, and then categorized by a trained scorer into one of the six "cognitive-developmental stages" which Kohlberg (1984) had defined. In the early version of the MJI, the scorer was to read through the complete answers of a participant and then decide which "stage" of moral reasoning would best fit them.

Kohlberg based his method on two assumptions or postulates, namely that people's moral cognitions are organized as a structural whole and that they develop in a pre-determined invariant sequence. He considered the scoring to be valid only if it agrees with these two postulates. Since the interview data did not agree well enough with these two postulates, he and his students revised the scoring system several times in order make it better fit with the data (Kohlberg 1984).

Inevitably, Kohlberg's measurement model came under attack from objective psychologists. These questioned the MJI's "objectivity" and "reliability" (Kurtines & Greif 1974). They argued that morality must be assessed (1) Through judging their behavior by external standards of morality (instead of talking to them about their behavior and other people's behavior), and (2) By scoring small pieces of behavior instead of looking at it holistically.

In response to these attacks, he and his collaborators also changed the method of scoring in order to make it more agreeable with the requirements of objective psychology (Colby *et al.* 1987). But they not only changed the method but actually jettisoned Kohlberg's original concept of moral competence:

• While Kohlberg (1958, 1963) originally defined moral behavior as behavior which is based on *internal* moral principals, the MJI uses *external* standards for scoring the individuals' responses: "I include in my approach a normative component. (...) That is, I assumed the need to define philosophically the entity we study, moral judgment, and to give a philosophic rationale for why a higher stage is a better stage" (Kohlberg 1984, 400; see also Lind & Nowak 2015).

• Originally Kohlberg based the Stage scores on a holistic analysis of people's total response pattern. He considered *structure* as the hallmark of his cognitive-structural approach: "The structure to which I refer is (...) a construct to subsume a variety of different manifest responses. The responses of subjects to the dilemmas and their subsequent responses to clinical probing are taken to reflect, exhibit, or manifest

the structure" (Kohlberg 1984, 407). Later he and his colleagues fragmented the interviewees' response into "items." They instructed the scorer to score each item individually instead: "Each item must have face validity in representing the stage as defined by the theory" (Kohlberg 1984, 410). However, through this itemizing of the scoring method, the cognitive-structural properties of an individual's response pattern become invisible. To reclaim some of their original structural idea, they instructed the scorer to put an answer into a higher stage, if it was "included in a higher stage idea." They argue that "ideas are often expressed within a higher stage context, which when taken literally or out of context would be scored at a lower stage" (Colby *et al.* 1987, 177). This attempt to save their original cognitive-structural feature of the method came under attack by objective psychologists again (Rest 1979, 60).

• Whereas Kohlberg formerly defined moral competence as an *ability* and tested this ability by confronting the respondents with tough probing questions and counterarguments, later he and his students eliminated these tasks in the revised MJI, presumably for the same reasons as for itemizing the scoring; namely to increase the statistical "reliability" of the MJI: "Test reliability and test construct validity are one and the same thing" (Kohlberg 1984, 424).

Similarly, objective tests of moral development like the *Defining Issues Test* (DIT) by Rest (1979), the *Socio-Moral Reflection Measure* (SRM) by Gibbs *et al.* (1992), and Haidt's (2007) test of moral values even take this accommodation of the definition of moral competence to standardized testing a step further: They score their respondents' answers in regard to *external* standards. Moreover, while some claim that they assess the *structural* properties of people's moral judgment, their scoring methods chop up the person's structure of judgments into atomized items, thus making structural information invisible. Some even claim to measure moral *competence*, but their tests lack any moral task.

You can't have your cake and eat it. Kohlberg agreed to improve the "reliability" of his clinical interview method at the expense of its theoretical validity (Lind 1989). This means that he actually gave up his original concept of moral competence in order to comply with the doubtful psychological assumptions underlying the so-called objective approach.

The "Objective" Approach

"Objective" psychologists believe that psychological theories bias measurement and that we better do without them. Measurement, they demand, should be based only on visible acts or behavior, but not on theory (Watson 1923).

However, they cannot in fact avoid theoretical assumptions. Instead of psychological theory, they base their measurement on *statistical theory*. This theory determines which

items and which scoring methods are regarded as valid. In other words, their statistical models define their object of measurement. Statistical models, they seem to believe, are more objective than psychological models. But they are not, as we will see.

The famous *Studies in the Nature of Character by* Hartshorne and May (1928) are a good example of objective behaviorists' approach to psychological measurement. Funded by a church organization, they wanted to test experimentally the hypothesis that character exists and that it is fostered through religious instruction. They confronted participants with situations in which they were tempted to cheat and observed how they reacted. They recorded the agreement or disagreement of these reactions with their standard of honesty. They explicitly discarded any psychological and philosophical interpretations of their subjects' behavior, because "no progress [of psychological science] can be made, however, unless the overt act be observed and, if possible, measured without any reference, for the moment, to its motives and its rightness or wrongness" (Hartshorne & May 1928, 11).

Obviously, the authors believed that we are able to read the character strength of a person directly from his or her reactions to temptations to deceive, like reading a temperature scale: the current temperature is simply the reading on the display plus/minus some error of reading or of malfunctioning of the scale. Similarly, they believe that we can reduce measurement error just by reading those reactions several times and calculating the average score.

However, objective measurement is also based on a theory, not on psychological theory but on a statistical theory such as "Classical Test Theory" or the *Theory* of *Mental Tests* (Gulliksen 1950), and its variants like "Item Response *Theory*," IRT. Notably, their theory is not about psychological objects but about statistical constructs, for example, about "latent variables," "latent classes," or statistical "factors." Through this theory, they create their own object of measurement, which may best be described as a "*homo statisticus*." Although the textbooks on CTT and ITR are usually voluminous (e.g., Cohen & Swerdlik 2012, has 612 pages), this homo statisticus is described by a very simple statistical formula: Y = T + e. This formula means that the reading of the scale ("Y") is simply the addition of the subjects' "true" behavior ("T") and some random error ("e"). The formulas of more sophisticated statistical test theories are more complex but are essentially based on similar statistical assumptions (Wilson 2005).

Can this *homo statisticus* be used to understand, predict and enhance human behavior? Can we use this statistical construct, for example, for examining the empirical validity of psychological theories of intelligence and morality, or for judging the efficacy of therapeutic and educational programs, or for evaluating students' achievement? The answer is no. This becomes obvious when we translate the statistical formulas underlying this construct into plain language. They allege that:

• *Observation is simple*. Objective psychologists believe that we can directly read the participants' behavior without any psychological interpretation. As we have seen above, they believe that, for example, the participants' behavior in an honesty experiment

enables us to directly read his or her character. By definition this behavior is not affected by any other factor like the type of temptation, the participants understanding of the test, or by their moral competence. To use another example, if a person gets 105 points on an IQ test, by definition it means nothing other than that this person really has an intelligence of 105 plus/minus some random error of the test.[2] No other interpretation is considered.

• *Error is random*. Any aberration of this statistical model from the real data is believed to be caused only by a *random* error of "reading the scale," meaning that no systematic factor of the participant or of the testing circumstances affects our reading of human behavior.

• *Repeated observation of identical behavior is possible*. Because objective psychologists believe that any error is purely random, it averages to zero. Therefore, they assert, measurement error can be simply reduced to any smallness simply by repeating the reading as often as needed (the so called "law of large numbers"). But this requires us to believe that people respond to replications of the questions or task always in the same way, and that they are willing to do so. But not even objective psychologists seem to believe this. They hardly ever confront participants with identical questions or identical tasks. They reason that people would remember their answers, or refuse repetitions. So even behaviorists admit that there are internal factors (like remembering, thinking, vigilance) interfering with observation and that, therefore, variation of behavior should not be regarded just as random error.

• *Similarity of behavior can be determined purely statistically*. Because objective psychologists avoid psychological concepts, they use statistical means for defining the "similarity" of tasks and questions. They define two behaviors as similar if the participants show them together. So, for example, if people answer two different questions in the same way, they are considered similar, or, if they solve task A and also task B in a math test, these two tasks are considered similar. If the items do not show statistical similarity, they are excluded from the test even though they may be considered as highly valid by experts on the subject matter. Note that if all test items which threaten the reliability of the test are excluded from the test as "dissimilar," the measurement model becomes immune to refutation through data. This immunization violates a basic standard of good science, namely refutability (Popper 1967). It also calls the objectivity of objective psychology into question and creates an illusionary reality. For example, Burton (1963) argued that the studies by Hartshorne and May (1928) would have actually proven the existence of a uniform character if the researchers had eliminated all experiments from their analysis that were "unreliable." In other words, Burton reasons that there are two groups of people: those who are always honest and those who are always dishonest, and never in between, in all thinkable situations – *except* in all those situations in which they

2 Variants of the CTT like Items-Response-Theory are more complex but rest basically on the same idea, namely that the behavior results from a random process (Allerup 2007).

behave differently.

• *Error and reliability are an attribute of the measurement instrument*. If they were an attribute of the measurement, they would not change from one application to another. But they do. Item selection does NOT lead to a stable estimate of a test's reliability, but it varies from one test sample to another and from one test administration to another. For example, even though PISA tests are carefully trimmed on the basis of many prior studies and the replacement of "unreliable" tasks, the final tests still deviate substantially from the statistical model on which their construction was based (Wuttke 2006; Jablonka 2007). If data change, it is not because of the tests. Tests are mostly, if not always, perfectly stable. Just observe a printed copy of the PISA-test for some months: you will find no change!

Objective psychologists like to compare their tests to the measurement by craftsmen and astronomers. Carpenters usually read their meter stick twice. This is enough to make sure that they do not accidently saw the beams for a house construction too long or too short. The one-time repetition has the advantage that it hardly affects its object (although their yard stick may leave some marks behind) and that the interval between the two readings is so short that the object does not change during the repetition. Observing human behavior is much trickier. Do we really always read the same thing when we repeat our observations like a carpenter does? In certain contexts, it may suffice to repeat a test question only once to make sure that it is correctly recorded. But in contrast to the carpenter's wood, people try to make sense out of test questions. So people may feel annoyed when being asked the same questions twice without a cause. For example, if we ask a person twice how she feels, she will answer the second question only if we explain that we did not understand her first answer, or that we wanted to observe change. But in the latter case, a different answer does not indicate an error but a change of feeling. In these cases, the repetition of the test question does not produce random error, but rather a systematic change of behavior.

Astronomers repeat their measurements more often. They do this because they want greater precision than a carpenter. Since many observations require a longer period, their targeted star (or the Earth) may move in the meantime, and their data reflect not only random measurement error but also a change of location. This will bias their measurement and the repetitions do not average to zero. Astronomers can differentiate such systematic influences from random reading error by looking at the distribution of their data. Only in as far as their data are distributed like a bell do they consider them to be caused only by reading error.

In contrast, objective psychologists usually avoid testing the hypothesis of random error and thus they overlook any systematic bias and ambiguity of their measurements (Wuttke 2007; Jablonka 2007). They may overlook, for example, as Scott (1968) showed, that scores in the middle range of an attitudes scale can have three very different meanings:

They can mean, as researchers mostly assume, (1) that the respondents have a medium attitude toward the declared object of the scale (like "conservatism"). But these scores can also mean (2) that they do not have such an attitude at all, but instead rate the items in regard to other criteria. Or these scores could mean (3) that the respondents have a differentiated attitude which involves more than the one attitude.

In order to clarify this ambiguity, I re-analyzed the findings from a longitudinal study on university students' political attitudes (Lind 1985a). The authors of this study reported that at in the first semester, students' attitudes become more liberal, and after graduation they become more conservative again. They interpreted these changes of students' attitudes as a consequence of their adaptation to different environments, which presumably changed from conservative to liberal (university) and than back again to conservative (workplace).

However, my secondary analysis of their statistics for "measurement error" over the span of university study, actually revealed a *structural* transformation of students' political attitudes: first the error was large, then decreased and then increased again. This supports the hypothesis that students had hardly any "conservative" attitude when they entered university. Then they developed a consistent (liberal) attitude, and finally their attitude became more differentiated, so that "measurement error" increases again and the scores moved back again to the middle of the conservatism scale. In other words, the students did not just adapt to their environment but also developed a higher competence for political reasoning. It was the researcher's statistical model which made students' structural development look like a pure "to and fro" of attitudes.

The blindness of objective psychology to structural aspects of human behavior explains Hartshorne and May's (1928) failure to produce evidence for the existence of character. Only after the completion of their study did Hartshorne and his colleagues admit that excluding internal traits from their observations was a mistake: "The essence of the act is its pretense. Hence [character] can be described and understood only in terms of the human elements in the situation. It is not the act that constitutes the deception, nor the particular intention of the actor, but the relation of this act to his intentions and to the intentions of his associates" (Hartshorne & May 1928, 377) The authors also admitted the blindness of their measurement model to the competence aspect of character: "A trait such as honesty or dishonesty is an achievement like ability in arithmetic, depending of course on native capacities of various kinds" (Hartshorne & May 1928, 379). Already some years earlier, the psychiatrist Levy-Suhl (1912) was surprised to find in his study of juvenile delinquents that they upheld the same moral values as non-delinquent youth. Therefore he hypothesized that they actually must differ in respect to their moral maturity, which psychologists were not able to measure at that time.

Another example for the discrepancy between the statistical measurement model and psychological reality is the OECD's *Programme for International Student Assessment* (PISA). The physicist Joachim Wuttke (2007) found much "evidence for multidimensiona-

lity," that is, for the existence for several internal factors. This contradicts the measurement model on which the PISA tests are based. The evidence, he notes,

> is even more striking on the background that the cognitive items actually used in PISA have been preselected for unidimensionality: Submissions from participating countries were streamlined by 'professional item writers,' reviewed by national 'subject matter experts,', tested with students in think aloud interviews, tested in a pre-pilot study in a few countries, tested in a field trial in most participant countries, rated by expert groups, and selected by the consortium (…). Only one-third of the items that had reached the field trial were finally used in the main test. Items that did not fit into the idea that competence can be measured in a culturally neutral way on a one-dimensional scale were simply eliminated. Field test results remain unpublished, although one could imagine an open-ended analysis providing valuable insight into the diversity of education outcomes. This adds to Olsen's (…) observation that in PISA-like studies the major portion of information is thrown away (Wuttke 2007, 249–250).

If objective psychologists used this thrown-away information they could interpret respondents' test scores more adequately. They would discover, for example, that the same task which is designed to challenge the respondents' *math competence* might actually challenge quite different dispositions, namely their ability to guess the "right" answer, their ability to copy it from other test takers, their knowledge of how to handle tests (test skill), their ability to stay awake on long testing cycles, and their ability to master their test anxiety, just to name a few of the factors which can influence a testee score. Or they might discover that "wrong" answers do not indicate a lack of math competence but that the testee made only a small error, or was not able to read the often wordy instructions quickly enough or was blocked by test anxiety. (Wuttke 2007) Similarly, behaviorist psychologists, who operationally define participants' moral character as an "honest" reaction to a situation of temptation, give them a high score regardless of whether or not these actually have high moral standards, or only incidentally acted "honestly" in this situation, or succeeded without the need to cheat because they knew all the answers (in fact cheating correlated negatively with IQ), or wanted to help a friend by letting her copy their test answers. So these scores have highly ambiguous psychological meaning.

Calling all these possible causes of test scores "random error" prevents any improvement of these tests and any progress of psychology as a science (Rosenthal & Rosnow 1997; Loevinger 1957). Moreover, it also undermines the trust in the validity of these tests. How can we expect consumers to trust tests, when even the chosen "test format or method of assessment can cause large differences in student scores?" (Walberg *et al.* 1994, 232) How can we rely on expensive studies like PISA for educational policy-making if it "is dominated and driven by psychometric [i.e., statistical] concerns, and much less by educational," writes the nuclear physicist Sjoberg (2007, 212).

How can we call these tests "psychometric" if there is no "psycho" in its metric? While the physical units of measurement are physically defined and standardized, the units of "standardized psychometric tests" are not defined psychologically and are not standardized objectively but only statistically. Their metric changes with the data of each

study, like rubber bands which stretch and bend as needed but are not *reliable* in the true sense of this word.

In spite of their blindness in regard to psychological theories, objective "psychologists" claim that their statistical models can be used to evaluate psychological theories, therapeutic methods, educational policies and competencies of people. They underpin this claim with a naming trick: (a) They give their statistical constructs psychological names like intelligence, character, or conservatism, and (b) They equate pattern of correlations across groups of people with an individual mind's "structure." But like family names, these names do not actually establish a real relationship between statistics and psychology. Or would Mrs. Miller allow an unrelated Mr. Miller to share her bedroom, just because he bears the same family name?

Anyway, this trick seems to work. World-wide, millions of dollars are spent every year on "objective" tests of academic abilities, vocational skills, character, mental disorders, and so on, in the hope that they can help to improve therapy, education and the politics of mental health and education. These tests have severe consequences for millions of students, job applicants, career seekers, mentally ill people, teachers, educational policy makers and many more who are tested many times throughout their lives, and also for decision-makers who base their policies on reported test scores. Because these tests measure something different from what they pretend to measure, they can cause a lot of damage. If these tests are bad, they will mislead us when we use them to evaluate methods and policies of therapy and education. If, for example, bad teaching practice produces higher scores on these tests than good teaching practice, they will defeat our educational system (Sjoberg 2017).

The dilemma of objective psychologists, it seems, is rooted in the ambiguous meaning of the word "objective." This word can take on quite different meanings:

• *Transparency*: This is an essential requirement of real science and good psychological practice. Only if data collection and scoring are fully transparent and uniform can they be critically examined by third parties. The questions and tasks of objective tests are usually transparent but often not available for the independent experts. The scoring of the answers is obscure for the customers. Instead of reporting the numbers of solved tasks, the scores are multiplied to make differences look large, and are transformed to make them look like a bell-shape. Ironically, the bell-shape indicates that the scores are pure error scores. Natural traits are hardly ever distributed like that: "An investigation of the distributional characteristics of 440 large-sample achievement and psychometric measures found all to he significantly non-normal at the alpha .01 significance level" (Micceri 1989, 156; see also Walberg *et al.* 1984). Finally, test scores are often obscure because important information like item selection and participants' attrition rates is held back.

• *Freedom from theory*: To be objective we need an object. Theories are an essential

basis of any measurement. If we want to measure a psychological object like an orientation or a competence, we need a psychological theory to define its nature. If we ban any psychological theory from our measurement method, we deprive it literally of its object. The test's reliability and precision become meaningless and its results useless.

• *Statistical standards*. The term "standardized" in the word "standardized tests" is actually a misnomer, because it does not mean that an individuals' test scores depend on a fixed standard but it means that an individual's scores depends on other people's test scores, namely how they compare to the scores of some sample of people. Such relative "standards" suggest wrong interpretations. For example, if a student solves ten more tasks on a test than he did last time, and if at the same time the members of the standardization sample also solve ten more tasks, his score (e.g., "percentile") will not increase and thus make him look as if he did not learn anything.

• *External or internal standards*. Objectivity is often used to mean external standards for scoring the participants' responses. However, objectivity can also require that we score a test in regard to the individuals' own standards, for example, if we want to measure how much progress they have made in regard to their *own* learning aims, or if we want to measure moral competence, which is defined as behaving in accordance with one's *own* moral principles.

How to Make Psychological Traits Visible

As I have shown, both mainstream approaches to measuring psychological attitudes and competencies are questionable. It has been often suggested to ease these problems by combining them in educational research and evaluation. But combining two bad meals does not make a good dinner. We should rather seek to find a better way of measurement which can replace the currently used ones.

As we have seen above, objective psychologists assume that *internal* psychological traits are not directly observable, and that structure is irrelevant, that is, that individual responses to test questions are unambiguously revealing the human trait under investigation and no other ones. In contrast, subjective psychologists target internal psychological traits can be made visible only by subjective methods on subjective ratings instead of on direct observations. But, as Jean Piaget (1965) admits, this is not a solution: "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 (...), reflection (...) or psittacism (...). We do not claim to have solved the problem completely. Only direct observation can settle it" (Piaget 1965, 115).

Already a hundred and fifty years ago the Dutch psychologist Franciscus Donders (1969/1868) showed how we can *directly* observe psychological traits. He was probably the first who discovered that we can test measurement hypotheses in the same way as we test hypotheses about the impact of external factors on human behavior. He hypothesized

that humans are not merely machines who always react to stimuli like an automaton, but that they also *think* when it is needed. To test this hypothesis he designed a simple experiment for which he constructed an ingenious time recorder for measuring very short reaction times. When he gave his participants clearly distinct stimuli, they reacted as quickly as an automaton. But when he gave them similar stimuli, they presented them with a "dilemma", so that their reactions took much longer. Obviously under the second condition they had to *think* before reacting.

The problem of the ambiguity of human behavior has been solved in principle by the Hungarian-American psychologist Egon Brunswik (1955). He has shown how we can disentangle multiple factors of behavior with what he called the "diacritical method." His idea was that we must design our observation as a multivariate experiment, in which the traits that are believed to determine particular responses are used as "design factors." Only such a structural experimental design of our observation can make the determining traits.

On the basis of Donders and Brunswik's ideas, I have developed the concept of *Experimental Questionnaire*, EQ (Lind 1982; 2019). EQs make human traits directly visible without involving dubious assumptions. They do not require statistical expertise in order to see the trait under investigation. The translation of the visible results into numerical scores is only done in order to facilitate the statistical analysis of mass data.

EQs confront the participants with a carefully designed *pattern* of stimuli, tasks, questions, or situations. The pattern is designed as an individual *multivariate experiment*. The design-factors of this multivariate experiment are chosen to directly correspond with the dispositional factors that are hypothetically involved in the participant's response to those tasks, questions or situations. That is, the construction of the design-factors of an EQ requires a psychological theory about the measurement object. When the design-factors of EQs are chosen to be independent of each other, we can literally see the impact of each hypothesized factor on a participant's responses in the pattern of an individual's responses, in a similar way as we can read brain activities from the monitor of a brain scanner.

On the base of this new methodology, I constructed the first objective *Moral Competence Test*, MCT (formerly called *Moral Judgment Test)* (Lind 1978, 2019). After reviewing a vast amount of research (Lind 1985b; 2002: Lind & Nowak 2015) and considering modern ethical theories (especially, Habermas 1990), I have defined moral competence *as the ability to solve problems and conflicts on the basis of moral principles through thinking and discussion instead of though violence, deceit, or complying with others* (Lind 2019). Like Kohlberg (1963) I consider as criterion for moral competence inner moral orientations instead of external standards. Moral competence, as the MCT defines it, is the ability to behave in accordance with one's own moral principles instead of with conformity to other people's judgments.

More specifically, moral competence becomes visible when a person judges the

arguments concerning a controversial decision in regard to the arguments' *perceived moral quality* instead of their *opinion agreement*. As Keasey (1974) observed in a series of experiments, this ability seems to be low in most people. So we decided to use this task for measuring moral competence. If people have not developed such a moral sense, they cannot solve problems und conflicts through moral thinking and moral discussion but must use violence, deceit or submission to others.

In order to make visible people's ability to *rate arguments in regard to their moral quality instead of to their opinion-agreement or to the particular context*, we used three hypothesized traits as design-factors for the MCT:

(1) *Dilemma context*: The participants are confronted with two stories in which a protagonist has to make a difficult decision. They are asked to take sides: Was the protagonist's decision right or wrong?

(2) *Opinion-agreement*: After each story the participants are to rate several arguments supporting and opposing their own opinion. They should say how much they reject or accept them on a scale from -4 to +4).

(3) *Moral quality*: All the arguments have all been painstakingly written so as to represent a clearly distinct *moral quality*, namely one of the *six types of moral reasoning* described by Kohlberg (1984). In order to secure their *theoretical validity*, the arguments were reviewed by several experts of Kohlberg's stage-typology and then revised accordingly (Lind 1978; Lind & Wakenhut 2010).

Thus each item of the MCT represents a specific manifestation of the three dispositional factors which may determine people's judgment. The items of the MCT have a 6 x 2 x 2 multivariate experimental design. Due to this experimental design, we can literally see the respondent's degree of moral competence directly by looking at their *pattern* of responses. Only the whole pattern of a respondent's behavior contains the structural information which defines moral competence. If we looked only at isolated responses, we would not be able to see their structure. Isolated items are bare of any structure.

In order to facilitate further statistical analysis, one can translate this visible pattern into the numerical C-score (C for competence). The C-score is the proportion of individual judgment variation caused by the moral quality of the arguments as compared to the respondent's total judgment variation. The C-score ranges from 0 to 100, the higher scores indicating higher moral competence. Mean C-scores are rarely higher than 30, indicating that for most people it is indeed rather difficult to engage in a moral discourse. Most people judge arguments mostly, or even solely, on the basis of their agreement with their opinion.

The measurement theory on which the MCT is based can be rigorously tested without saving circularity. As already mentioned, the content validity of its items (arguments) has been examined through ratings by several experts of Kohlberg's typology

of moral orientations. The structural validity of the MCT can be experimentally tested in regard to four prominent hypotheses of cognitive-developmental theory:

• *Competence nature of morality*: In contrast to many other psychologists, Piaget (1965) and Kohlberg (1958, 1964) hypothesized that moral behavior is not only affective in nature but also cognitive, that is, it is not only determined by people's moral orientations (values, attitudes, principle, and so on) but also by their moral competence. This competence hypothesis has been clearly supported by experiments. While participants can be instructed to fake their moral orientations upward (Emler *et al.* 1983), the same kind of instruction fails to make participants fake their MCT's C-scores upward (Lind 2002). Experiments also showed that the ability to estimate other people's moral competence is positively correlated with their own moral competence (Wasel cited in Lind 2002).

• *Moral competence is a unique skill*. The ability to solve moral dilemmas is not just a linguistic skill but is a unique competence. This has been shown by the research team of Kristin Prehn (2013) of the Charité Hospital in Berlin. Moral competence as measured with the MCT correlates highly with neural activities in the right dorso-lateral prefrontal cortex (DLPC) when subjects' brain activities are studied in a brain scanner while they are confronted with moral problems: the lower their C-score is the longer their right DLPC is busy. This phenomenon does not show when the subjects are confronted with linguistic problems.

• *Hierarchical preference order*: Kohlberg (1958, 1984) and Rest (1969) hypothesized that the six types of moral orientations – which were identified on the basis of philosophical analysis – form a universal order of moral adequacy. This hypothesis lets us predict that people will prefer these types according to their order. This, as Karl Popper (1968) would say, is a very informative, because daring, hypothesis. Since the six types of orientation can be ordered in 720 (= 6!) different ways, the risk of a coincidental confirmation of the hypothesis is very small ($p = 1/720 = 0.0014$). Note that this risk is much smaller than the risk of accidently confirming a conventional statistical hypothesis ($p < 0.05$). The risk of accidental confirmation becomes *extremely* small if, for example, we test this hypothesis with ten people ($p = 0.0014^{10}$). It is even more astonishing that his hypothesis has been almost unanimously confirmed in many empirical studies (Lind 1986; 2002).

• *Simplex structure of moral orientations*: Kohlberg (1958) hypothesized that the correlations between the six types of moral orientation show a "simplex structure," which means that neighboring orientations correlate more highly with each other than with more distant orientations. The many MCT studies support this hypothesis with very few exceptions (Lind 1978, 2002).

• *Affective-cognitive parallelism*: Piaget (1976) hypothesized that affective and cognitive aspects of behavior are "parallel." This hypothesis has two important implications. First, Piaget saw orientations and competences not as separable components, but as two *distinguishable aspects of behavior*. This means that he rejected the prevailing notion

that human traits are components which can be separated from each other and from behavior, and can be measured separately. So all attempts to assess them separately are in vain. Second, his parallelism hypothesis lets us predict that the higher people's moral competence is, the more clearly they will prefer higher types of moral orientations, and reject low, inadequate types. Only with the MCT does it become possible to test Piaget's hypothesis, because only this test allows us to measure affect and cognition as distinct but inseparable aspects. So far, all MCT studies have very clearly supported Piaget's parallelism hypothesis (Lind 2002, 2013).

The exceptionally clear confirmation of these four core hypotheses shows that moral competence is something real and than it can be made visible in an objective and valid way. This gives us the opportunity to test hypotheses about the relevance, development and teachability of moral competence in an objective and unbiased way.

- *Relevance:* Already studies using Kohlberg's *Moral Judgment Interview* have found that moral competence determines our social behavior more than any other psychological trait. Experimental and correlation studies using the MCT confirm and extend these findings. Moral competence seems to be highly instrumental for such important behaviors like helping people in distress, engaging in democracy, obeying the law, respecting a contract, blowing the whistle, fulfilling academic achievement requirements, and making quick decisions (Lind 2019). A C-score above 20 seems to be critical. Only when people have a moral competence higher than 20 does their behavior in experiments show some determination by inner moral orientations. People who lack any moral competence either conform to the perceived opinion of the majority of people or to the orders of an authority, like in the Milgram experiment (Kohlberg 1984).

- *Development:* MCT research has refuted the cognitive-developmental postulate of invariant sequence of development: People's moral competence can regress if they do not have an opportunity to use it for a longer period of time (Lind 2000; Schillinger 2006; Lupu 2009; Saeidi 2011). Yet it supports the findings of Kohlberg and his associates that moral competence can be effectively fostered through certain methods of dilemma discussions (Lind 2002).

- *Education:* Finally, the MCT lets us objectively and economically measure the efficacy and efficiency of methods and programs of moral education, like the *Konstanz Method of Dilemma Discussion* (KMDD) (Lind 2002, 2019; Hemmerling 2014).

Conclusion

Experimentally designed tests let us make psychological traits visible, validly and objectively. Experimentally designed tests accomplish what subjective psychologists always wanted to achieve, and what objective psychologists could not deliver: They make it possible to measure the properties of humans' internal traits through direct observation of their behavior. Moreover, experimentally designed tests allow us to examine the

truth of the assumptions on which they are based. Experimentally designed tests can be used not only in moral psychology but in any field of psychology in which testable theories and clear definitions of their objects are available.

A final caveat: experimental questionnaires should be used only for research and evaluation of methods and programs, not for evaluating people. That is, the MCT must not be used for high-stakes testing of students, teachers, or named institutions. There is no evidence that test-based sanctions lead to better learning and better behavior. Moreover, sanctions undermine the validity of psychological tests and, therewith, impede their usefulness for improving therapy and education (Amrein & Berliner 2002; Ravitch 2013; Koretz 2017). When used for high-stakes testing, tests wear out within a few years and must be substituted by new content. Thus their findings can be compared only through daring statistical constructions (Linn 2010). In contrast, the MCT celebrates its 44th anniversary and still has not had to be changed (beside a few minor editorial corrections). Thus it has provided us with a great wealth of data that stretch over a long period of time and across many cultures. This in turn allows us to test many hypotheses on the nature, relevance, and teachability of moral competence.

Acknowledgements

I would like to thank the following persons for encouraging me to write this article and for commenting on at least parts of earlier drafts of it: Małgorzata Steć, Hans Brügelmann, Jim Fearns, Wilhelm Kempf, Martina Reinicke, and Aiden Sisler.

References

Allerup P. 2007. "Identification of Group Differences Using PISA Scales – Considering Effects of Inhomogeneous Items," in S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA* (pp. 175–202). Berlin: LIT-Verlag.

Amrein A. L. & Berliner D. 2002. "High-stakes Testing, Uncertainty, and Student Learning," *Education Policy Analysis Archives* 10 (18):1–74.

Ausch R. 2016. *Methodological Problems with the Academic Sources of Popular Psychology*. Lanham: Lexington Books.

Bailey K. 1994. *Method of Social Research*. New York: The Free Press.

Block J. 1967. "Advancing the Psychology of Personality," in D. Magnusson & N. Endler (Eds.), *Personality at the Crossroads* (pp. 37–63). Hillsdale, NJ: Lawrence Erlbaum.

Bühler K. 1927. *Die Krise der Psychologie* [The crisis of psychology]. Jena: Gustav Fischer.

Cohen R. J. & Swerdlik M. E. 2018. *Psychological Testing and Assessment. An Introduction to Tests and Measurement.* 9th edition. New York: McGraw Hill.

Colby A., Kohlberg L., Colby A., Speicher B., Hewer A., Candee O., Gibbs J., & Power C. 1987. *The Measurement of Moral Judgment: Standard Issue Scoring Manual*, Vol. I & II. New York: Cambridge University Press

Donders F. C. 1969 (Orig. 1868). "On the Speed of Mental Processes," *Acta Psychologica. Attention and Performance II*, 30:412–431. Retrieved from http://www2.psychology.uiowa.edu/faculty/mordkoff/InfoProc/pdfs/Donders 01868.pdf (on February 25, 2021). DOI: 10.1016/0001-6918(69)90065-1

Emler N., Renwick S., & Malone B. 1983. "The Relationship between Moral Reasoning and Political Orientation," *Journal of Personality and Social Psychology* 45:1073–80.

Graumann C. F. 1960. "Eigenschaften als Problem der Persönlichkeitsforschung," in P. Lersch & H. Thomae (Eds.), *Handbuch der Psychologie, Bd. IV, Persönlichkeitsforschung und Persönlichkeitstheorie* (pp. 87–154). Göttingen: Hogrefe.

Gregory R. J. 2018. *Psychological Testing: History, Principles, and Applications*. London, England: Pearson Publisher.

Gulliksen H. 1950. *Theory of Mental Tests*. New York: Wiley.

Habermas J. 1990. *Moral Consciousness and Communicative Action*. Cambridge: MIT Press.

Haidt J. 2007. "The New Synthesis in Moral Psychology," *Science* 316:998–1002.

Hartshorne H. & May M. A. 1928. *Studies in the Nature of Character. Vol. I: Studies in Deceit, Book I & II*. New York: Macmillan.

Hemmerling K. 2014. *Morality behind Bars – An Intervention Study on Fostering Moral Competence of Prisoners As a New Approach to Social Rehabilitation*. Frankfurt/Main: Peter Lang.

Jablonka E. 2007. "Mathematical Literacy: Die Verflüchtigung eines ambitionierten Testkonstrukts in bedeutungslosen PISA-Punkten," in T. Jahnke & W. Meyerhöfer (Eds.), *Pisa & Co. Kritik eines Programms* (pp. 247–280). Hildesheim: Franzbecker.

Keasey Ch. B. 1974. "The Influence of Opinion-agreement and Qualitative Supportive Reasoning in the Evaluation of Moral Judgments," *Journal of Personality and Social Psychology* 30:477–482.

Keller M. 1990. "Zur Entwicklung moralischer Reflexion: Eine Kritik und Rekonzeptualisierung der Stufen des präkonventionellen moralischen Urteils in der Theorie von Kohlberg," in M. Knopf & W. Schneider (Eds.), *Entwicklung. Allgemeine Verläufe – Individuelle Unterschiede* (pp. 19–44). Göttingen: Verlag für Psychologie.

Kempf W. 1981. *Testtheorie.* University of Konstanz, Dept. of Economy and Statistics, a mimeographed paper.

Kohlberg L. 1958. *The Development of Modes of Moral Thinking and Choice in the Years 10 to 16*. University of Chicago. Unpublished doctoral dissertation.

Kohlberg L. 1964. "Development of Moral Character and Moral Ideology," in M. L. Hoffman & L. W. Hoffman (Eds.), *Review of Child Development Research*, Vol. I (pp. 381–431). New York: Russel Sage Foundation.

Kohlberg L. 1976. "Moral Stages and Moralization: The Cognitive-developmental Approach," in T. Lickona (Ed.), *Moral Development and Behavior: Theory, Research and Social Issues* (pp. 31–53). New York: Holt, Rinehart & Winston.

Kohlberg L. 1979. *The Meaning and Measurement of Moral Development. The Heinz Werner Lecture Series,* Vol. 13. Worcester, MA: Clark University Press.

Kohlberg L. 1984. *The Psychology of Moral Development.* Vol. II*: Essays on Moral Development*. San Francisco: Harper & Row.

Koretz D. 2017. *The Testing Charade. Pretending to Make Schools Better*. Chicago: University of Chicago Press.

Kurtines W. M. & Greif E. B. 1974. „The Development of Moral Thought: Review and Evaluation of Kohlberg's Approach," *Psychological Bulletin* 81:453–470.

Levy-Suhl M. 1912. „Die Prüfung der sittlichen Reife jugendlicher Angeklagter und die Reformvorschläge zum § 56 des deutschen Strafgesetzbuches," *Zeitschrift für Psychotherapie und Medizinische Psychologie* 4:232–254.

Lind G. 1978. "Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts," in G. Portele (Ed.), *Sozialisation und Moral* (pp. 171–201). Weinheim: Beltz.

Lind G. 1982. "Experimental Questionnaires: A New Approach to Personality Research," in A. Kossakowski & K. Obuchowski (Eds.), *Progress in Psychology of Personality* (pp. 132–144). Amsterdam, NL: North-Holland.

Lind G. 1985a. "Attitude Change or Cognitive-moral Development? How to Conceive of Socialization at the University," in G. Lind, H. A. Hartmann, & R. Wakenhut (Eds.), *Moral Development and the Social Environment. Studies in the Psychology and Philosophy of Moral Judgment and Education* (pp. 173–192). Chicago: Precedent Publishing.

Lind G. 1985b. "The Theory of Moral-cognitive Development: A Socio-psychological Assessment," in G. Lind, H. A. Hartmann, & R. Wakenhut (Eds.), *Moral Development and the Social Environment* (pp. 21–53). Chicago: Precedent Publishing Inc.

Lind G. 1986. "Cultural Differences in Moral Judgment Competence? A Study of West and East European University Students," *Behavior Science Research 20:*208–225.

Lind G. 1989. "Essay Review: 'The Measurement of Moral Judgment' by Anne Colby, Lawrence Kohlberg *et al*.," *Human Development 32:*388–397.

Lind G. 2000. "Moral Regression in Medical Students and Their Learning Environment," *Revista Brasileira de Educacao Médica* 24(3):24–33.

Lind G. 2002. *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung*. 2nd Edition. Berlin: Logos-Verlag.

Lind G. 2005. *The Cross-cultural Validity of the Moral Judgment Test:* Findings from 28 Cross-cultural Studies*. Presentation at the conference of the American Psychological Association, August 18–21, 2005, Washington, D.C.

Lind G. 2013. "Thirty Years of the Moral Judgment Test:* Support for the Dual-aspect Theory of Moral Development," in C. S. Hutz & L. K. de Souza (Eds.), *Estudos e pesquisas em psicologia do desenvolvimento e da personalidade: uma homenagem a Angela Biaggio* (pp. 143–170). São Paulo: Casa do Psicólogo.

Lind G. 2017a. "From Donder's Dilemma to Objective Internal Assessment: How Experimental Developmental Psychology Can Contribute to Moral Education," *Psychologia Rozwojowa* 22(3):15–24.

Lind G. 2017b. *Wie kann man moralische Orientierung im Verhalten erkennen? Von Brunswiks diakritischer Methode zum Moralische Kompetenz-Test (MKT)*. Beitrag zur Moralforschertagung, January 26–28, 2017, Universität Leipzig.

Lind G. 2019. *How to Teach Morality.* Berlin: Logos.

Lind G. & Nowak E. 2015. "Kohlberg's Unnoticed Dilemma – The External Assessment of Internal Moral Competence?," in B. Zizek, D. Garz, & E. Nowak (Eds.), *Kohlberg Revisited* (pp. 139–154). Rotterdam – Taipei – Boston: Sense Publishers.

Lind G. & Wakenhut R. 2010. "Testing for Moral Judgment Competence," in G. Lind, H. A. Hartmann, & R. Wakenhut (Eds.), *Moral Judgment and Social Education* (pp. 79–105). Rutgers, NJ: Transaction Books, 2nd edition.

Linn R. L. 2000. "Assessments and Accountability," *Educational Researcher 29(*2):4–16.

Loevinger J. 1957. "Objective Tests as Instruments of Psychological Theory," *Psychological Reports* 3:635–694.

Lupu I. 2009. „Moral, Lernumwelt und Religiosität. Die Entwicklung moralischer Urteilsfähigkeit bei Studierenden in Rumänien in Abhängigkeit von Verantwortungsübernahme und Religiosität." Doctoral dissertation, University of Konstanz, Germany.

Meehl P. E. 1958. "When to Use Your Head Instead of the Formula?," in H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science* (pp. 498–506). Minneapolis, MI: University of Minnesota Press.

Micceri T. 198. "The Unicorn, the Normal Curve, and Other Improbable Creatures," *Psychological Bulletin* 105(1):156–166.

Piaget J. 1965 (Orig. 1932). The *Moral Judgment of the Child*. New York: The Free Press.

Piaget J. 1976. "The Affective Unconscious and the Cognitive Unconscious," in B. Inhelder & H. H. Chipman (Eds.), *Piaget and His School* (pp. 63–71). New York: Springer.

Popper K. R. 1968. *Conjectures and Refutations*. New York: Harper & Row.

Prehn K. 2013. "Moral Judgment Competence: A Re-evaluation of the Dual-Aspect Theory

Based on Recent Neuroscientific Research," in E. Nowak, D. Schrader, & B. Zizek (Eds.), *Educating Competencies for Democracy* (pp. 9–22). Frankfurt am Main – Bern – Brussells – New York: Peter Lang Editions.

Ravitch D. 2013. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Book.

Rest J. R. 1969. "Level of Moral Development As a Determinant of Preference and Comprehension of Moral Judgments Made by Others," *Journal of Personality* 37(1):220–228.

Rest J. R. 1979. *Development in Judging Moral Issues.* Minneapolis, MI: University of Minnesota Press.

Rosnow R. L. & Rosenthal R. 1997. *People Studying People. Artifacts and Effects on Behavioral Research*. New York: Freeman & Co.

Ryle G. 1949. *The Concept of Mind*, Chicago: University of Chicago Press.

Scott W. A. 1968. "Attitude Measurement," in G. Lindsey & E. Aronson (Eds.), *Handbook of Social Psychology*. Vol. II (pp. 204–273). Reading, MA: Addison-Wesley.

Saeidi-Parvaneh S. 2011. *Moral, Bildung und Religion im Iran – Zur Bedeutung universitärer Bildung für die Entwicklung moralischer Urteils- und Diskursfähigkeit in einem religiös geprägten Land.* Doctoral Dissertation, University of Konstanz.

Schillinger M. 2006. *Learning Environments and Moral Development: How University Education Fosters Moral Judgment Competence in Brazil and two German-speaking Countries*. Aachen: Shaker Verlag.

Schoenfeld A. H. 1999. "Looking Toward the 21th Century: Challenges of Educational Theory and Practice," *Educational Researcher* 28:4–14.

Sjoberg S. 2007. "PISA and 'Real Life Challenge': Mission Impossible?," in S. T. Hopfmann *et al.* (Eds.), *PISA zufolge PISA* (pp. 201–224). Berlin: LIT-Verlag.

Sjoberg S. 2017. "PISA Testing – A Global Educational Race?," *Europhysics News.* https://www.europhysicsnews.org/articles/epn/pdf/2017/04/epn2017484p17.pdf

Travers R. M. 1951. "Rational Hypotheses in the Construction of Tests," *Educational and Psychological Measurement* 11:128–137.

Walberg H. J., Strykowski B. F., Rovai E., & Hung S. 1984. „Exceptional Performance," *Review of Educational Research* 54(1):87–112.

Walberg H. J. & Haertel G. D. 1994. „The Implications of Cognitive Psychology for Measuring Student Achievement," in OECD (Ed.), *Making Education Count* (pp. 219–236). Paris: OECD.

Watson J. B. 1970 (Orig. 1924). *Behaviorism*. New York: Norton.

Wilson M. 2005. *Constructing Measures. An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum Associates Publishers.

Wittgenstein L. 1953. *Philosophical Investigations*. Trans. by G.E.M. Anscombe. New York: The Macmillan Company.

Wuttke J. 2005. “Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung” [Errors, biases, uncertainties in PISA interpretation], in T. Jahnke & W. Meyerhöfer (Eds.), *Pisa & Co. Kritik eines Programms* (pp. 101–154). Hildesheim: Franzbecker.

Wuttke J. 2007. “Uncertainties and Bias in PISA,” in S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA* (pp. 241–264). Berlin: LIT Verlag.

(for more references on EQ and MCT see: http://moralcompetence.net)

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #4, pp. 39-58. DOI:10.14746/eip.2021.1.4

Moral Competence and Foundations of Moral Judgment. An Empirical Exploration of Concepts and New Possibilities for Understanding

Bogdan Popoveniuc

(University Stefan cel Mare of Suceava/West University of Timişoara; Romania; bpopoveniuc@usm.ro)
ORCID: 0000-0002-4615-7594

Abstract: The present study examines the relation between the moral intuitions proposed by the 'Moral Foundations Theory' according to J. Haidt and the moral orientations in relation with moral competence, as described by the 'Dual-Aspect Model of Moral Behaviour' according to G. Lind.

It is an empirical exploration of the relation between the five foundational domains (and/or the corresponding two higher-order clusters) and the moral orientations, as this results from the theoretical assumptions of both theories, and the influence on and of the moral competence on these aforementioned relations. The shared affective and, at the same time, habitual characteristics of moral intuitions and moral orientations, although based on different theories (intuitionism, respectively, cognitivism) should converge in a dynamic relationship.

The basic motivation is the need to overcome the affective-cognitive disjunction in the study of moral judgment, which is artificial. The human (moral) reasoning is always a situated and situating one, in a symbolic narrative universe, in which the cognitive-affective dichotomy forms a dynamic unity. This overstated dichotomy is only a methodological necessity, but unfortunately hypostasized in an explanatory mechanism that leads to the impermissible simplification of the phenomenon that it seeks to understand.

The empirical results confirm the limits of both cognitive and intuitions paradigms and advocates for a more nuanced and dynamic approach in understanding the moral judgment within a more comprehensive conception on human personality. Finally, the implications for the contemporary psychology' paradigm are discussed.

Keywords: moral competence; moral judgment; moral intuitions; moral orientations; axiology.

1. Introduction

The study of the mechanisms of moral judgment is both a challenging and a difficult task, as the moral functioning involves the cooperation of complex interrelated components. Even the most used and praised cognitive theories on moral judgment are based on more comprehensive psychological constructs, which exceeds by far what is

usually understood by cognition. The neo-Kohlbergian approach is based on moral sensitivity – involving perception, sensitivity and interpretation of situations –, the moral judgment – related to reasoning, judgment and reflection –, the moral motivation –encompassing motivation and concentration –, and the moral character – linked to initiation and completion of action (Rest 1983).

The constraints imposed by the contemporary research paradigm, methodological limitations and practical aspects make it difficult to verify complex models involving emotional, cognitive, value-based and motivational aspects, which involve self-narrative building upon a complex fabric of cultural and ideological meanings. For this reason, studies on moral judgments are often limited either to highlighting particular components of them, or to studying the manifestation of moral behavior in a particular context.

For example, the dominant paradigm in the study of reasoning, judgments, and decision-making takes for granted the (conflicting) relationship between affectivity and reason. The main line of approach towards moral judgments opposes the emotions to reason, and defines the methods of investigation. Depending on the stimulus, the proposed prototypical situations highlight one or the other of these dimensions, as they are framed within the experiment task (Ellemers, van der Toorn, Paunov, & van Leeuwen 2019).

> The different models of morality that have appeared in the literature over the years may be a direct consequence of the different moral situations considered by the researchers who have proposed them: observe humans as they try to solve complex moral dilemmas, and you are likely to propose a model of morality that relies heavily on high-level reasoning; ask them how they feel about disgusting immoral acts, and you are likely to conclude that morality is all about gut reactions that require little rational deliberation (Monin, Pizarro, & Beer 2007b, 223; see also Monin, Pizarro, & Beer 2007a).

In other terms, the research within the "moral domain" is threatened by the effect of the self-fulfilling prophecy phenomenon (Jussim 1986). Studies on the functioning of the moral judgment are limited to particular relationships and interactions within these complex models. In terms of values, motivations, emotions or attitudes, the problem of researching the role of affectivity in relation to moral reasoning and moral judgment has certain intrinsic methodological limitations.

Most of the studies do not set out to check this dichotomy from the beginning, and, hence, limit *ab initio* the possibility of identifying more nuanced relationships. Even thought, they seem to justify, rather, a perspective based on the idea of continuity between "intuition" and "reasoning" as "two facets of the same process which spans from fast, immediate, and certain answer to slow, conscious and elaborate judgments" (Dellantonio & Job 2012, 241).

The present study aims to highlight possible relationships between *intuitive aspect* (mostly affective, pre- or subconscious, and also involving automatic reasoning) and *cognitive aspect* (which are mostly based on post-hoc conscious reasoning) of the moral judgment. The relationship between affectivity (understood as moral intuition) and

reason – or the moral reasoning – is not necessarily either disjunctive, or contradictory. The gap between theoretical approaches is so big, that it seems they are dealing with different phenomena. These heterogeneous perspectives compose a dismantled imagine of human moral psychology. The picture of human moral is painted in flashy colors of the clashing self-subsistent monolithic and conflicting mechanisms (cognitive, affective, motivational etc.). Greene's theory on moral judgment is reflecting this state of affairs, by starkly opposing emotion (through the lens of a deontological-prone judgment) to cognition (utilitarian-prone judgments), as "competing subsystems in the brain" (Green *et al.* 2004). The images on moral judgment are inevitably half-done, as the two main theories (also used in this study) – the Moral Foundations Theory (MFT) and the Cognitive Development Theory (CDT) – highlight two partial images on moral judgment:

MFT	event	→	intuition	→	judgment	↔	post hoc reasoning
CDT	event	→	?	→	ante hoc reasoning	→	judgment

But, if the assumptions of the two models are merged, it can get presumably a better image of the moral reasoning, which seems to suggest the following logical flow:

event → [interpretation (automatic ante hoc reasoning)] → intuitive judgment (affective aspect) → [post hoc reasoning (motivated or reflective)] → moral judgment (cognitive aspect).

2. Theoretical Background of the Research

2.1 The Moral Foundations Theory (MFT)

The Moral Foundations Theory (Haidt 2001, 2013; Haidt & Graham 2007; Graham *et al.* 2011) proposes a nativist, intuitionist and pluralistic perspective on morality. It starts from the evolutionary premise that there is a nativist pre-organization of the human mind that facilitates the acquisition of behaviors, norms and values, which are involved in adapting to a set of recurring social problems. They have a vital function of controlling and suppressing natural selfishness in order to make social life possible. The five main foundations of morality, based on quick and automatic moral intuitions, are those related to "harm/care," "fairness/reciprocity," "ingroup/loyalty," "authority/respect," and "purity/sanctity". The first two foundations collapse in personal-individualizing category, as they are oriented on the rights and liberties of all individuals, while the last three in group-binding, focusing on the group cohesion). Beside these, the authors have also identified other possible candidate dichotomies such as "liberty/oppression," "efficiency/waste," and "ownership/theft" (Graham *et al*. 2013). I use also the three clusters version of collapsing the moral foundations as suggested by the original source of MFT Shweder's (1997) "big three" of morality: autonomy ("harm" & "fairness), community ("loyalty"

& "authority"), and divinity ("spirituality"), because the fifth foundation Spirituality is heavily loaded with religious significance

The relevance of this theory for the present study is that it highlights the active role of the sets of values (rooted in the same moral foundations) to which the individual adheres to, and in relation to the way he or she reasons. In the Moral Foundations Questionnaire, the authors use two sets of items. The first set includes relevance-items, which measure personal theories (self-theory) about the moral judgment (the perceptions that an individual's has about his own moral values), the second set are judgment-items, concrete examples about which individual makes moral judgments.

The relevance-items are formulated abstractly by reference to the group – generally family or nation – in order to avoid cultural conflict on certain sensitive issues, e.g., gay rights or the right to bear arms. Judgment scales were added to the relevance ones in order to minimize the impact of the variation; this was based on the set of answers – e.g., a person can consider all aspects as morally relevant – as well as to contextualize the abstract items. These items also have the function to balance the differences between the explicit self-theories related to how one makes moral judgments with the effective moral judgment (Nisbett & Wilson 1977).

2.2 Dual-Aspect Theory of Moral Behavior (DATMB)

Dual-Aspect Theory of Moral Development is rooted in Kohlberg's cognitive theory of moral development. It is based on the principle that moral competence is "the capacity to make decisions and judgments which are moral" (i.e., based on internal principles) and to act in accordance with such judgments" (Kohlberg 1964, 425). Therefore, one might argue that morality cannot be reduced to principles, attitudes, values, intuitions, moral stances, but all of these are involved more or less in the moral judgement. However, in order to reflect individual's free-will and deliberate decision, without which the moral responsibility is a *flatus vocis*, the structure or the cognitive scheme is essential to underpin the moral orientation. At the same time, affectivity (or emotion) does not come as a complementary, opposite or distinct aspect of cognition, but should be conceived as the secondary aspect of the moral behavior. Within the theory of the dual aspect of moral judgment, proposed by Georg Lind, moral judgment (cognitive dimension) and moral orientation (affective dimension) are two *aspects*, but not *components*, of the moral behavior. Moral competence is nothing but the ability to use consistently and differently a certain moral orientation for making moral judgments in different social situations (Lind 2016). In other words, the "virtue" of self-consistent moral behavior or the verticality (read, consistency) of moral character. The more moral I am, the more consistent I am in my judgments (and conduct), in various situations, with different actors and regarding different deeds.

In the research literature, MFT and CDT are two different paradigms, which reflect different psychic phenomena. Whilst MFT aims at the range of social concerns grouped

semantically (linguistically) around the term "moral," CDT aims to capture the end-point of a process of developing increasingly differentiated thinking about social issues. "On offer, then, are *not* two functionally equivalent conceptions of the moral domain, but two constructs that are doing different theoretical work in very different theoretical frameworks" (Maxwell & Beaulac 2013, 378). However, both theories share the same assumption, i.e., that affectivity has a well-defined position in moral judgments and this study tries to see if they converge in this point, as they should.

3. Literature Review

The viability of my hypothesis was suggested by the few (and) scattered studies on this topic. The analysis of the literature suggested the possibility of the existence of relations between certain foundations and moral orientations, in relation with the moral competence. The link between the moral foundations and the moral competence seems to be the most documented, mainly by the negative relationship discovered in several studies between social conservatism and the moral competence (measured with either DIT2 or MCT: [Emler 2002; Candee 1976; Fishkin, Kenniston, & MacKinnon 1973; Lind, Sandberger, & Bargel 1985; Nassi, Abramowitz, & Youmans 1983; Raaijmakers, Verbogt, & Volleberg 1998; Simpson 1987]).

On the other hand, moral competence was found to negatively correlate with the "binding" foundations (Graham, Haidt & Nosek 2009). However, comparing the populations of America and Israel, Gross (1996) has shown that these differences are considerably tempered when education and socioeconomic status are taken into account, although the samples are relatively small, four groups of about 50 subjects.

The analysis of the research carried out on this topic led to the identification of just three articles that directly address the issue of the relationship between moral competence and moral foundations. Investigating the relationship between psychopathy and moral orientation, Gay, Vitacco, Hackney, Beussink and Lilienfeld (2018) find a positive correlation of moral competence with the foundation of "fairness/reciprocity" and a negative one with "ingroup/loyalty" in one study. However, the second study finds no correlation and the third study finds only a negative correlation between moral competence and the two foundations of "binding" foundations (Gay, Lishner, Vitacco, & Beussink 2019).

The study made by Trups-Kalne and Dimdins (2017) seems to be the only one that aims to test directly the relationship between moral foundations and moral orientations, and moral competence respectively. obtaining only a negative relationship between moral competence and the "binding" foundation. The study also finds a positive correlation between the arguments corresponding to conventional stage 3 of Kohlberg's model – based on the appeal on solidarity and group cohesion – and the importance given to "binding" foundations. Paradoxically, a negative correlation was noticed between moral

competence and "fairness". Their conclusion was that the rational-cognitive and social-intuitionist perspectives on moral judgment use different conceptual constructions, the only correlations obtained being opposed, overall, to the theoretical predictions.

The present study intends to answer this challenge and to test this connection between the two sets of constructs – i.e., the moral foundations and the moral orientations, respectively, the moral competence, articulated within two alternative theoretical paradigms which investigate the territory of moral judgments/decisions. It aims to identify the inter-relations between the aspects of relevance (evaluative, i.e., moral values) and the criteria of judgment (moral orientations), which are both related to the capacity of moral reasoning, conceived as the level of moral competence.

The first hypothesis is that the moral competence correlates positively with the "individualising" of moral foundations ("harm/care," "fairness/reciprocity") and, negatively, with "binding" ones ("ingroup/loyalty," "authority/respect," and "purity/sanctity"). Given that the MCT uses judgment items, the correlations of moral competence with judgment subscale should be stronger than that on the relevance subscale.

The second hypothesis is that moral intuitions is related with the moral orientations, a relationship moderated by the level of moral competence as it follows:

(H2.1) *The moral orientations corresponding to the conventional stages (3, 4) are in a positive relationship with the "biding" foundations*;

(H2.2) *The moral orientations corresponding to the postconventional stages (5, 6) are in a positive relation with the foundations of "individuality*";

(H2.3) *The correlations between moral orientations and the moral foundations as measured on the judgment subscale are weaker than the measures on the relevance subscale*;

(H.2.4) *The predicted patterns of correlations are stronger for higher moral competence*;

The present study has an exploratory purpose and tone, which is justified partly by the lack of substantial investigation of these relations within the present literature.

4. Methodology of the Research

4.1 Instruments

Many studies have tried to clarify the relationship between values and moral judgment using various instruments, i.e., Kohlberg´s Moral Judgment Interview (Kohlberg 1981, 1984) or Defining Issues Test (Rest, Narvaez, Thoma, & Bebeau 1999), and Schwartz´ Personal Values Questionnaire (SPV) (Schwartz *et al*. 2012) or Rokeach Value Survey (RVS) (Rokeach 1973) respectively. So far, the results from these studies

proved inconclusive, with some studies showing no influence at all (Gay *et al*. 2018), or displaying weak relations (Ostini & Ellerman 1997), or some partial relations (Helkama 1982; Diessner, Mayton, & Dolen 1993; Lan, Gowing, Rieger, McMahon, & King 2010). Other studies have argued for the existence of the mediating function of values – for example, between personality traits, or empathy, and ethical competence (Pohling *et al.* 2016), whilst others revealed a more systematic relationship (Lan, Gowing, McMahon, Rieger, & Friz 2008).

Moral foundations were measured using the Moral Foundations Questionnaire (MFQ) proposed by Jesse Graham and Jonathan Haidt, which was based on the MFT (Graham *et al*. 2011). Moral orientations and moral competence were measured using "Moral Competence Test" (MCT), an instrument developed throughout a 40 years period by Professor Georg Lind, based on a "Dual-Aspect Theory of Morality and Moral Development" (Lind 2016).

Although the instruments of measurement used in these studies are not seen as being convergent, their shared purpose and complementarity made them useful for constructing the research hypotheses, e.g., the consistency discovered between Schwartz's Personal Values Questionnaire (SPV) (Schwartz *et al*. 2012) and Haidt's dimensions predict unique variance in morality attitudes, behavior, and individual-differences (Feldman 2020). For the purpose of my research, I consider that both of these two instruments have managed to instill the necessary confidence, which is so important when aspiring to have reliable measurements. (In the very recent literature, however, appears studies that question the viability of Moral Foundation instrument). The confidence was directly provided by their background theoretical assumptions, which considered the intertwined relationship between emotion and cognition as key, and how these instruments themselves were constructed to allow the observation and consideration of these both components. I have therefore chosen these two instruments due of their properties and operational value, which make them compatible for the study of the interaction between values (intuitions) and moral judgment.

The MFQ contains two sets of questions: i.e., one based on the assessment of the relevance of the five types of foundations and the other being based on the assessment of judgments made according to them. On the other hand, the MCT aims at measuring the moral competence index, which is based on the assessment of agreement (consistency of answers) with arguments corresponding to the six types of moral orientations inspired by Kohlberg's stage model. MCT also aims at identifying the moral orientation preferred by the respondent.

Therefore, although the two instruments are rooted in distinct theoretical traditions, they both meet at the level of judgment and in relation to the relevance given to moral values. The instruments are complementary in analyzing the image of the interaction between values (intuitions) and judgments, respectively, moral reasoning, although in different ways and on different theoretical backgrounds. At the same time, I have tried

to investigate whether there was a convergent validity between the moral foundations and the moral orientations. According to the theories on which the instruments are based on, both constructs share the assumption of the innate-cultivated hybrid character of moral judgment. If modular innateness is salient in the case of the MFT, the modular and gradually constructed structure of moral orientations from the theory of CDT could be naturally promoted by such preparedness. This assumption if (indirectly) tested, can be an argument for the convergence of the two theoretical perspectives.

I choose MCT instead of DIT2, because the two instruments measure the affective and cognitive aspects in different ways, although in an integrated or even mixed way. DIT2 measures how consistently (cognitive aspect) the subject prefers (affective aspects of) post-conventional arguments. What is difficult to determine in the case of DIT2 is whether this consistency is due to moral development or simply the desire to defend a certain point of view, because DIT2 is related to the preference of reasoning according to a particular stage (Bataglia, de Morais, & Lepre 2010). By contrast, the advantage of MCT in measuring the coherence of answers is that it assesses the capacity for moral reasoning independently of the subject's orientation (Lind 2016), i.e., it measures the ability to manage the task of differentiating the quality of arguments regardless of the preference for one action or another (Ishida, 2006).

For both instruments, the official Romanian translations indicated by their respective authors were used. In the case of MFQ, the version translated by Oana Luiza Rebega and Irina Pitică, Livia Apostol (retranslation) from the https://moralfoundations.org/questionnaires/ was applied, and for MCT it was used the Lind-Chicu (2004) version, provided by its author.

4.2 Research design and procedure

The study was exploratory, transversal, comparative – i.e., gender, level of education – and correlational – i.e., factors from MFQ and MCT, while the selection of the sample was non-probabilistic.

The two questionnaires were applied on a number of 982 students during class hours, under supervision, during 2019. The questionnaires were applied anonymously (coded) by their teachers, as a practical application of theories thought on classes. The participants have received activity points for their voluntary participation, and were informed by further use of the data. The data was analyzed according to authors' instructions: those respondents who failed the two dummy questions for MFQ and those with less than one missing answer for MCT were excluded. In order to ensure a better relevance of the data that could have been strongly distorted by the age difference, the answers of the middle school students were not retained for the analysis, these being used later for the validation of MCT and comparative analyses. The final database included 739 subjects: 509 (68.87%) female, 230 (31.12%) male, age 14 to 54 years (94% 15 to 25 years, M = 19.34, SD = 5,232), education level: 452 (61.25%) high school students, 287 (38.8%) university students.

5. Research Results

For MFQ, the results of Romanian students were significantly higher scores than those of American original study in the evaluation of moral foundations, both globally and on the two subscales, especially on "biding" foundation. Female subjects higher average scores than male subjects on all moral foundations (have statistically significant, excepting "in-group/loyalty" foundation), and their scores were higher also on both sub-scales. At the same time, the subjects have higher score on judgment subscale than on relevance subscales all foundations. Comparing the scores of high-school and university students, the same tendency was noticed, the university students having significantly higher scores on all moral foundations.

These results were consistent with those obtained for moral orientation scores assessed by MCT: the high school students had statistically significant lower scores than that of the university student population. An unexpected result was the statistically significant lower score at moral competence for the university student population than for the high school students, the average difference being 3.8 points ($M_{highschool}$ = 22.0, SD = 15.03 compared to $M_{university}$ = 18.2, SD = 12.98). In other words, university students have higher moral orientations and higher scores on all moral foundations, but lower levels of moral competence.

In the case of female subjects, the difference of moral competence C-index is even more statistically significant from high school to university, with an average decrease of 4.3 points from 21.8 to 17.5, from high school to university, compared to only 2.9 points for male subjects (from 22.5 to 19.6).

According to the initial hypothesis, the analysis of the correlations between the moral orientation and the moral foundations (Kendall's tau_b two-tailed) shows a slight positive correlation with the "individualizing" foundations (τ = .048, p = .042, 1-tailed). In addition, those with moral orientations corresponding to stages 1 and 2 are receptive only to the foundations of the "community" ("loyalty" and "authority"), those from level 3 onwards being receptive to all 5 moral foundations (Table 1).

	O M.O.	Moral orientations					
		Orient 1	Orient 2	Orient 3	Orient 4	Orient 5	Orient 6
Moral foundation		*5 factors*					
Care	.028	.018	-.004	.065*	.075**	.090**	.121**
Fairness	.042	.019	.013	.083**	.067*	.095**	.139**
Loyalty	-.005	.062*	.064*	.111**	.100**	.090**	.112**
Authority	.001	.102**	.067*	.075**	.060*	.051	.077**
Purity	-.008	.035	.075**	.073**	.073**	.069**	.089**
		2 factors					
Individualizing	.048	.023	.004	.085**	.077**	.106**	.151**
Biding	-.009	.079**	.088**	.102**	.094**	.093**	.112**

Progressivism	.059*	-.058*	-.088**	-.027	-.024	.034	.013
		3 factors					
Autonomy	.048	.023	.004	.085**	.077**	.106**	.151**
Community	-.005	.096**	.078**	.106**	.093**	.081**	.105**
Divinity	-.008	.035	.075**	.073**	.073**	.069**	.089**

Kendall's tau_b two-tailed [1]

** Correlation is significant at the .05 level (2-tailed)*
*** Correlation is significant at the .01 level (2-tailed)*
Note: M.O. – moral orientation

Table 1: Correlations between Moral Foundations and Moral Orientations.

This observed correlations' pattern is more obvious in terms of relevance subscales than the judgment ones in the case of the five foundations (except for "purity"). The preference for "autonomy" foundations is consistent with increasing moral orientation, whilst those in moral stages 3 and 4, corresponding to conventional level of moral orientation, are mostly receptive to "solidarity" foundations.

The analysis of gender differences shows a much greater consistency in the use of the five moral foundations for female subjects (in particular, to female high school subjects, i.e., higher moral competence), compared to male subjects where this consistency has not been manifested more than in the case of the last two moral orientations.

This result is supported by the education level analysis where the consistency in the use of all five moral foundations is found in high school students (with higher moral competence score), but not in university students (lower moral competence score).

In order to have a better image of the correlations between the moral foundations and the moral level the moral orientations was grouped on the three original foundations on which they were built, i.e. the preconventional, the conventional and the postconventional level. The analysis of the correlations with the moral foundations highlights the fact that the preconventional level does not correlate with the person-centered moral foundations ("harm" and "fairness,") instead they are more oriented towards the foundation of "authority" and "loyalty". Conventionally-oriented people are also more receptive to the aspects of "solidarity," while those who prefer post-conventional orientations are receptive to both sets of values ("individualization" and "solidarity") in a greater extent, to the fundamentals of the latter ("harm" and "fairness"). Table 2 shows the correlations between moral foundations and the level of moral orientation.

		Moral orientation level		
	Moral level	Preconventional	Conventional	Postconventional
Moral foundations	*5 factors*			
Care	.030	.011	.090**	.125**
Fairness	.047	.018	.089**	.142**

1 The Kendall's τb test was preferred over the Spearman's rank correlation ρ because in the literature there are opinions that claim that it would perform better on larger samples (see Zar 1996, 392).

Loyalty	-.010	.072**	.130**	.109**
Authority	-.004	.099**	.091**	.074**
Purity	-.007	.066*	.089**	.086**
	2 factors			
Individualizing	.051	.016	.100**	.154**
Biding	-.008	.098**	.124**	.113**
Progressivism	.074*	-.088**	-.034	.035
	3 factors			
Autonomy	.051	.016	.100**	.154**
Community	-.010	.101**	.127**	.103**
Divinity	-.007	.066*	.089**	.086**

Kendall's tau_b two-tailed

** Correlation is significant at the .05 level (2-tailed)*

*** Correlation is significant at the .01 level (2-tailed)*

Table 2: Correlations between Moral Foundations and Moral Level.

The correlations between moral foundations and the level of moral orientation are stronger in the case of those with a high level of moral competence than in the case of those with a low level of orientation, which is an indication that the relationship between intuitive foundations of moral judgment and moral orientations is influenced by moral competence. Table 3 compares the correlations between moral foundations and the level of moral orientation.

	C-index<20				C-index>20			
	M_{level}	Prec.	Conv.	Postc.	M_{level}	Prec.	Conv.	Postc.
Moral foundation					5 factors			
Care	,011	,012	,080*	,087*	,061	,015	,104**	,193**
Fairness	,034	-,001	,095**	,128**	,084†	,029	,091*	,194**
Loyalty	,017	,042	,084*	,073*	-,048	,111**	,190**	,169**
Authority	-,025	,046	,061	,055	,055	,142**	,142**	,153**
Purity	-,009	,053	,090**	,060†	,011	,054	,092*	,151**
					2 factors			
Individualizing	,030	,005	,091**	,113**	,092†	,025	,117**	,235**
Biding	-,004	,063†	,100**	,083*	,004	,125**	,159**	,187**
Progressivism	,079*	-,048	-,006	,049	,058	-,114**	-,068†	,005
					3 factors			
Autonomy	,030	,005	,091**	,113**	,092†	,025	,117**	,235**
Community	-,005	,056†	,087**	,077*	-,003	,144**	,189**	,173**
Spirituality	-,009	,053	,090**	,060†	,011	,054	,092*	,151**

Kendall's tau_b two-tailed

** Correlation is significant at the .05 level (2-tailed)*

*** Correlation is significant at the .01 level (2-tailed)*

† Correlation is significant at the 0,05 (1-tailed)

Note: MI_{evel} – moral level, Prec. – preconventional, Conv. – conventional, Postc. - postconventional

Table 3: The relationship between moral foundations and the level of moral orientation according to the level of moral competence.

Regarding moral competence, the analysis showed negative correlations with the foundation of "authority" and "purity" / "spirituality," "biding" and "community" and positive with "progressivism". These correlations, as was hypothesised, were due only to the scores on the judgment sub-scales (on relevance sub-scale no significant correlation was found), which were statistically significant along the lines of "authority", "purity", "solidarity", and "community".

For female subjects, only a positive correlation was found between moral competence and "progressivism" and a negative with "authority," and for male subjects – negative correlations with "authority" and "purity".

For high school students only "progressivism" correlated positively, but in a slight way, with the index of moral competence, and for university students there was a negative correlation with "purity".

Regarding the differences in the use of moral foundations according to the moral competence index (C-index), the analyses showed statistical differences only in the case of the "authority" foundation where those with low index appeal to a greater extent (M_{Clow} = 2.89 compared to M_{Chigh} = 2.78).

Regarding the level of education, the only statistically significant difference was the one related to the "purity" / "spirituality" foundation in the case students (M_{Clow} = 3.12, M_{Chigh} = 2.83).

The correlations between moral foundations and the level (low or high, cut-off point C-index = 20) of the moral competence index according to the level of moral orientation, it was also analyzed. It was found a relationship between the level of moral competence and the tendency of subjects – with conventional and postconventional moral orientations – to use more consistently all five moral foundations. The analysis of the correlation patterns on the two subscales (relevance and judgment) reveals that those with conventional and postconventional moral orientations and high index of moral competence display higher correlations between orientations and judgments, compared to those with low index of moral competence.

6. Discussion

Being mostly an exploratory study, extensive analyses were performed, beyond the initial hypotheses that were suggested by theoretical claims and the previous few studies. The aim was to make visible the more nuanced and subtle possible relations between the structure of moral intuitions and the moral orientations, in relation to the moral competence.

For the Social Intuitionist Model of moral judgment, moral intuitions are automated moral judgments, laden with affective valence (good-bad, like-dislike) and opposed to conscious and intentional process of reflective deliberation of moral reasoning (Haidt 2001, 818).

For the Dual Aspect Model of Moral Behavior, the moral orientations and the moral judgment are *aspects* of the moral behavior, not *components*, as suggested by Lind (2016, 51). Moral intuitions and the moral orientations seem to share both emotional roots and cognitive schema, which are unconscious and automated. Even though the trends of correlations were low, they were in line with the predicted second hypothesis. From the perspective of a "bijective" correlation, the first hypothesis is confirmed only partially. There were negative correlations between the level of moral competence and foundations and a negative correlation with those of "solidarity", "authority" and "purity". These correlations were due exclusively, as theoretically predicted, to the subscales of the judgment. This represents a supporting argument for the convergence of the measurements of the two instruments. On the other hand, there was no correlation between moral competence and the foundations of "individualisation" ("care" & "correctness") neither globally, nor on the two sub-scales (relevance, respectively judgment).

The second hypothesis was largely confirmed. Moral competence has a weak moderating effect on correlations between moral foundations and moral orientations, according with the predicted pattern. The moral orientations corresponding to the conventional level correlate most strongly with those of the "solidarity" foundations, and those corresponding to the postconventional level with the "individualisation" foundation. This is in agreement with other studies in the field; see, for example, Glover *et al.* (2014). The difference from previous studies was that correlations were identified at all stages corresponding to the conventional and postconventional moral level.

The moral competence correlates negatively, but only with the subscale of judgment, for the foundations of "authority" and "purity" of MFQ. As we have seen, the subscales of the "relevance" of MFQ is closer to the explicit reasoning within moral reasoning, expressing self-theory or second-order opinion about the criteria underlying one's own moral judgment or how people believe that moral judgments are, and because of this are affected by subjectivism and self-image. This would be an indication that maturation, unaccompanied by a development of moral competence, can lead to a dissociation between moral self-image and moral character.

On the other side, the results are promising because they suggest the possible inappropriateness of "bijective modular approach" on moral judgment mechanisms and, especially, of the relations between moral intuitions (emotion and unconscious reasoning) and moral reasoning.

The extensive analysis of the relations between the moral orientations and the moral foundations allowed exploring possible more complex relations than the "modular" ones (1 to 1 types), between these constructs within the moral judgment. The results support a new hypothesis, contrary to the one argued by Trups-Kalne & Dimdins (2017), who claimed that due to the increase in analytical complexity, an impairment of moral competence is manifested. My findings suggest that it seems that those with high moral competence are more able to operate consistently (principally) with all types of values/

intuitions and not vice versa. Indeed, the MFQ hypothesis, *by itself*, is not "an adequate measure to capture a more advanced moral functioning" as Glover *et al*. (2014) claims. It is not, also, representative at all for those at the preconventional and conventional stages of the development of moral judgment, as Maxwell and Beaulac (2013) argue.

It seems that the conclusions of the aforementioned studies are the result of "Procrustean bed" of moral psychological research, which neglect or diminish the relevance of semantic dimension of moral reasoning. One of the main shortcomings of mainstream cognitive theories on moral judgment is the dichotomy, in fact, antinomic perspective on conscious/unconscious, intuition/cognition, reason/emotion. The moral intuition is definitely more than automatic emotional reaction and comprise more cognition (heuristics) than most of the perspectives suggests. (Dubljević & Racine 2014) The other limitation is the negligible value assigned to the semantic aspects. The semantic intervenes at the level of reasoning (Stenning & van Lambalgen 2004) and emotional arousal or trigger.[2]

As it was proposed by Bucciarelli, Khemlani and Johnson-Laird (2008), the moral affective-cognitive hybrid functioning must not be so coherent and structured as the researchers are trying to make evident. Firstly, they said, there is no necessarily a single criterion of choosing moral propositions from a deontic set of moral principles. Secondly, the mechanisms underlying emotions and deontic evaluations could be very well independent and operate in parallel, and, consequently, some scenarios can elicit emotions prior to moral evaluations, with some other eliciting moral evaluations prior to emotions, and some eliciting them at the same time. Thirdly, the deontic evaluations seem to depend on inferences, either unconscious intuitions or conscious reasoning. And, forth, it is not necessary that a person's beliefs about what is, or isn't, moral to be either complete or consistent.

As Dellantonio and Job (2010) claim, the moral reasoning consists in specific operations not only with the concepts, but mostly on the features, from which these concepts are made of. There seems to be a semantic difference between the Externalized Semantics – conventional and the public dimensions of concepts, which are acquired through language and socialization, and are "characterised by rules whose aim is to assure the possibility of intersubjective communication" (Dellantonio & Job 2010, 507) and Internalized Semantics – or how information is used by people use (internally, i.e., in their mind) "to carry out categorizations and to understand the linguistic meaning". These two semantics come along with their own negative or positive value, and correspondingly

2 For example, in this research were a high rate of responses under cut-of value for the second dummy question of Moral Foundation Questionnaire ("Whether or not someone was good at math"). In the Romanian version, the term "right" from the questionnaire question "When you decide whether something is right or wrong…," translated as "correct" has its primary sense as "according to the rules in general", and its secondary one emphasis its moral meaning, which seems to be the other way around in comparison to the English meaning for "right". This became obvious when some subjects were questioned about their answer and they mentioned that they have thought it is important to have a developed mathematical, logical ability in order to make something right ("correct").

emotions, which can be similar or not.

7. Limitations

The Moral Competence Test passed all three psychological criteria of validity: (1) the six type of moral orientations were preferred in a increased order, (2) the parallelism between moral competence and type of moral orientations, and (3) the correlations among the preferences for the six types form a simple structure (Lind 2016). On the other side, the major limitation of the study was the low performance on the studied population of MFQ, which displayed only a low model fit. This could be the result of the authors' strategy for constructing the instrument by focusing on capturing different (theoretical possible) facets of each foundation with the risk of having dissimilar items that correlate moderately, as opposed to achieving high internal consistency (Graham *et al*. 2011). The very recent analyses increasingly reveal the necessity to amend and improve such instruments for measuring moral values particularly in intercultural research. A systematic content analysis of 539 studies (Tamul *et al*. 2020, in press) reveals that the mean Cronbach's alpha scores for four of the five subscales of Moral Foundations Questionnaire were below 70. Other studies on the 27 countries spanning the five largest continents found that it is difficult to replicate the five-factor model across a wide variety of populations (Iurino & Saucier 2020). My own analysis – which is in progress – does suggest that most of the studies take its psychometric properties for granted.

Other possible limitations could result from the application of the questionnaire during class hours. This could contribute to a certain "framing" of the data, given that the power dynamics in the classroom present certain particularities; the answers of the subjects are susceptible to be influenced by the context, the presence of the tenured teacher, the unexpected character of the task, etc.

8. Conclusions

There is a concordance, although not powerful, between moral foundations, as measured by MFQ, and moral orientations, assessed by MCT. At the same time, the level of moral competence acts as a moderating factor of this correlation between intuitions and orientations. Those people with higher moral competence, i.e., ability to judge consistently with a given set of moral principles, are able to employ more consistently their (educated) moral intuitions. The development of moral cognition is governed by higher schema and the concordance with triggered intuition is moderated by the general level of moral competence. Moral reasoning is not necessarily parallel with the moral intuition. People have the power to educate their moral intuitions, not only to construct moral judgments on top of them. Moral intuitions are rationally amenable and the patterns for automatic judgment foundation could be shaped by episodes of rational reflection (Sauer 2017).

Moreover, the study suggests a possible explanation for a possible theoretical outcome steaming from how the Moral Competence Test was designed, but which was not found in empirical data, i.e., the possibility of obtaining consistent judgments on low levels of moral orientation. The empirical research showed that this case is not common at all, but there was no explanation why this is happening. My study suggests that is possible that the individuals with low levels of moral orientation are unable to use consistent values/intuitions in their moral judgment. Their axiological reasoning is conjectural; they have no consistent moral system or perspective on their experience. They interpret and employ values/intuitions in their moral judgment according to their motivated reasoning and not as fixed fundamental criteria for judgment. They have no commitment for a certain type of values and judge using any intuitions at hand.

Finally, the results corroborate with those of Stenning and van Lambalgen suggesting that "«massive modularity» in cognition should be treated with some skepticism" (Stenning & van Lambalgen 2004, 523). The striving to discover innateness, substantial modularity and distinct mechanisms, in moral judgment and reasoning blow in the wind while the cognitive functioning could be more dynamic and hybrid than it is thought. The ability to judge according to higher stages of cognitive moral development does not mean that people will judge each time and/or exclusively in this manner regardless of the context, the experience, motivation, actors, and the particular situation, all influencing moral judgment strategies.

The meta-theoretical assumptions underlying the research in moral psychology are essential for the success of research programs. For reasons of convenience and congruence with methodological constraints of cognitive paradigm, the main research programs work predominantly with a narrow conception of the ethical domain divided between deontological versus utilitarian ethics. But there are other meta-ethical paradigms, as is the virtue ethics. This paradigm, consonant with the image of moral judgments as educated intuitions mentioned earlier, can be more suitable to explain human moral behavior, because this paradigm seems to be able to manage diverse situations with high ethical complexity without appealing to fixed specific rules. Of course, the virtue ethics is intrinsically related to personality, and it conceived as an organic whole and as an enterprise, which could mean that the very personality traits model would need to be enriched. Beside "dispositional traits" of well-established Big Five Model, the new one has to take into account, "individual's unique variation" on the general evolutionary design, "characteristic adaptations" and, most of all, "self-defining "life narratives", which are complexly and differentially situated in culture and social context (McAdams & Pals 2006). This can explain why moral values are hardly generalizable in various cultures, although theories are found to be universal. "The processes that underlie moral cognition may not be a human universal in any simple sense, because moral systems may play different roles in different cultures" (Sachdeva, Singh, & Medin 2011).

Only such comprehensive model seems to be able to accommodate appropriately the intuitions (rooted in the human evolutionary design), the moral orientations (derived from characteristic adaptation), the moral reasoning (influenced by self-narrative "complexly and differentially situated in culture and social context") in a more accurate image of what it is human moral psychology. The scientific paradigms are lenses for our eyes, which can re-present reality only in the way they are capable of. Therefore, there is a necessity to critically examine, not only the theory, instruments and methodology, but also the meta-theory which underlies the very research paradigm. Only in this way, scientific knowledge can provide a rich and appropriate image of the complex tri-unitary phenomenon which is the human being (Popoveniuc 2017).

References

Bataglia P. U. R., de Morais A., & Lepre R. M. 2010. "A teoria de Kohlberg sobre o desenvolvimento do raciocínio moral e os instrumentos de avaliação de juízo e competência moral em uso no Brasil" [Kohlberg's theory about moral judgment development and the instruments used for evaluation of moral judgment and moral competence in Brazil,] *Estudos de Psicologia* 15(1):25–32.

Bucciarelli M., Khemlani S., & Johnson-Laird P. N. 2008. "The Psychology of Moral reasoning," *Judgment and Decision Making* 3(2):121–139.

Candee D. 1976. "Structure and Choice in Moral Reasoning," *Journal of Personality and Social Psychology* 34:1293–1301.

Greene J. D., Nystrom L. E., Engell A. D., Darley J. M., & Cohen J. D. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44(2):389–400.

Dellantonio S. & Job R. 2010. "Morality According to a Cognitive Interpretation: A Semantic Model for Moral Behavior," in L. Magnani, W. Carnielli, & C. Pizzi (Eds.), *Model-Based Reasoning in Science and Technology. Studies in Computational Intelligence*, Vol. 314. Springer: Berlin – Heidelberg.

Dellantonio S. & Job R. 2012. "Moral Intuitions vs. Moral Reasoning. A Philosophical Analysis of the Explanatory Models Intuitionism Relies On," in L Magnani & L. Ping (Eds.), *Philosophy and Cognitive Science*. *Western & Eastern Studies* (pp. 239–262). Springer.

Diessner R., Mayton D., & Dolen M. A. 1993. "Values Hierarchies and Moral Reasoning," *The Journal of Social Psychology* 133(6):869–871.

Dubljević V. & Racine E. 2014. "The ADC of Moral Judgment: Opening the Black Box of Moral Intuitions with Heuristics About Agents, Deeds, and Consequences," *AJOB Neuroscience* 5(4):3–20.

Emler N. 2002. "Morality and Political Orientations: An Analysis of Their Relationship," *European Review of Social Psychology* 13:259–291.

Ellemers N., van der Toorn J., Paunov Y., & van Leeuwen T. 2019. "The Psychology of Morality: A Review and Analysis of Empirical Studies Published from 1940 through 2017," *Personality and Social Psychology Review* 23(4):332–366.

Feldman G. 2020. "Personal Values and Moral Foundations: Examining Relations and Joint Prediction of Moral Variables." (Unpublished manuscript, available at https://osf.io/6qs5g/, lact accessed on April 10, 2021).

Fishkin J., Keniston K., & McKinnon C. 1973. "Moral Reasoning and Political Ideology," *Journal of Personality and Social Psychology* 27(1):109–119.

Gay J., Vitacco M., Hackney A., Beussink C., & Lilienfeld S. 2018. "Relations among Psychopathy, Moral Competence, and Moral Intuitions in Student and Community Samples," *Legal and Criminological Psychology* 23*:*117–134.

Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S., & Ditto P. H. 2013. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism," *Advances in Experimental Social Psychology* 47:55–130.

Graham J., Haidt J., & Nosek B. A. 2009. "Liberals and Conservatives Rely on Different Dets of Moral Foundations," *Journal of Personality and Social Psychology* 96(5):1029–1046.

Graham J., Nosek B. A., Haidt J., Iyer R., Koleva S., & Ditto P. H. 2011. "Mapping the Moral Domain," *Journal of Personality and Social Psychology* 101:366–385.

Greene J. D., Nystrom L. E., Engell A. D., Darley J. M., & Cohen J. D. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44(2):389–400

Gross M. L. 1996. "Moral Reasoning and Ideological Affiliation: A Cross-national Study," *Political Psychology* 17(2):317–338.

Haidt J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108(4):814–834.

Haidt J. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. Pantheon Books: New York.

Haidt J. 2013. "Moral Psychology for the Twenty-first Century," *Journal of Moral Education* 42(3):281–297.

Haidt J. & Graham J. 2007. "When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals May Not Recognize," *Social Justice Research* 20:98–116.

Helkama K. 1982. "The Development of Moral Reasoning and Moral Values," *Acta Psychologica Fennica* 9:99–111.

Ishida C. 2006. "How do Scores of DIT and MJT Differ? A Critical Assessment of the Use of Alternative Moral Development Scales in Studies of Business Ethics," *Journal of Business Ethics* 67(1):63–74.

Iurino K. & Saucier G. 2020. "Testing Measurement Invariance of the Moral Foundations Questionnaire Across 27 Countries," *Assessment* 27(2):365–372.

Jussim L. 1986. "Self-fulfilling Prophecies. A Theoretical and Integrative Review," *Psychological Review* 93(4):429–445.

Kohlberg L. 1964. "Development of Moral Character and Moral Ideology," in M. L. Hoffman & L. W. Hoffman (Eds.), *Review of Child Development Research. Vol. I* (pp. 381–431). New York, NY: Russel Sage Foundation.

Kohlberg L. 1981. *Essays in Moral Development*. *Vol. I*, *The Philosophy of Moral Development*. New York: Harper & Row.

Kohlberg L. 1984. *Essays in Moral Development. Vol. II, The Psychology of Moral Development*. New York: Harper & Row.

Lan G., Gowing M., McMahon S., Fritz R., & Norman K. 2008. "A Study of the Relationship between Personal Values and Moral Reasoning of Undergraduate Business Students," *Journal of Business Ethics* 78(1-2):121–139.

Lan G., Gowing M., Rieger F., McMahon S., & King N. 2010. "Values, Value Types and Moral Reasoning of MBA Students," *Business Ethics: A European Review* 19(2):183–198.

Lind G. 2016. *How to Teach Moral Competence*. 1st Ed. Berlin: Logos Verlag.

Lind G., Sandberger J. U., & Bargel T. 1985. "Moral Competence and Democratic Personality," in G. Lind, H. A. Hartmann, & R. Wakenhut (Eds.), *Moral Development and the Social Environment* (pp. 55–78). New Jersey: Precedent Publishers.

Maxwell B. & Beaulac G. 2013. "The Concept of the Moral Domain in Moral Foundations Theory and Cognitive Developmental Theory: Horses for Courses?" *Journal of Moral Education* 42(3):360–382.

McAdams D. P. & Pals J. L. 2006. "A New Big Five. Fundamental Principles for an Integrative Science of Personality," *American Psychologist* 61(3):204–217.

Monin B., Pizarro D. A., & Beer J. S. 2007a. "Deciding Versus Reacting: Conceptions of Moral Judgment and the Reason-Affect Debate," *Review of General Psychology* 11(2):99–111.

Monin B., Pizarro, D. A., & Beer, J. S. 2007b. "Emotion and Reason in Moral Judgment: Different Prototypes Lead to Different Theories," in K. D. Vohs, R. F. Baumeister, & G. Loewenstein (Eds.), *Do Emotions Help or Hurt Decision Making?: A Hedgefoxian Perspective* (pp. 219–244). New York: Russell Sage Foundation Press.

Nassi A. J., Abramowitz S. I., & Youmans J. E. 1983. "Moral Development and Politics a Decade Later: A Replication and Extension," *Journal of Personality and Social Psychology* 45(5):1127–1135.

Nisbett R. E. & Wilson T. D. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84(3):231–259.

Ostini R. & Ellerman D. 1997. "Clarifying the Relationship between Values and Moral Judgment," *Psychological Reports* 81(2):691–702.

Pohling R., Bzdok D., Eigenstetter M., Stumpf S., & Strobel A. 2016. "What is Ethical Competence? The Role of Empathy, Personal Values, and the Five-factor Model of Personality in Ethical Decision-making," *Journal of Business Ethics* 137(3):449–474.

Popoveniuc B. 2017. "Evoluția psihologiei și structura trinitară a personalității" [The evolution of psychology and the trinitarian structure of personality], in *Filosofia* și *perspectiva umană*. Chișinău: UnAȘM, Biotehdesign.

Raaijmakers Q. A. W., Verbogt T. M. A., & Volleberg W. A. M. 1998. "Moral Reasoning and Political Beliefs of Dutch Adolescent and Young Adults," *Journal of Social Issues* 54(3):531–546.

Rest J. R. 1983. "Morality," in P. H. Mussen (Ed.), *Manual of Child Psychology.* 4th Ed., Vol. 3 (pp. 495–555). New York: Wiley.

Rest J. R., Narvaez D., Thoma S. J., & Bebeau M. J. 1999. "DIT2: Devising and Testing a Revised instrument of Moral Judgment," *Journal of Educational Psychology* 91(4):644–659.

Rokeach M. 1973. *The Nature of Human Values*. New York: The Free Press.

Sauer H. 2017. M*oral Judgments as Educated Intuitions: A Rationalist Theory of Moral Judgment*. MIT Press.

Sachdeva S., Singh P., & Medin D. 2011. "Culture and the Quest for Universal Principles in Moral Reasoning," *International Journal of Psychology* 46(3):161–176.

Schwartz S. H., Cieciuch J., Vecchione M., Davidov E., Fischer R., Beierlein C., Ramos A., Verkasalo M., Lönnqvist J.-E., Demirutku K., Dirilen-Gumus O., & Konty M. 2012. "Refining the Theory of Basic Individual Values," *Journal of Personality and Social Psychology* 103:663–688.

Shweder R. A., Much N. C., Mahapatra M., & Park L. 1997. "The 'Big Three' of Morality (Autonomy, Community, Divinity) and the 'Big Three' Explanations of Suffering," in A. M. Brandt & P. Rozin (Eds.), *Morality and Health* (pp. 119–169). Abingdom on Thames: Taylor & Frances/Routledge.

Simpson E. 1987. "The Development of Political Reasoning," *Human Development* 30(5):268–281.

Stenning K. & van Lambalgen M. 2004. "A Little Logic Goes a Long Way: Basing Experiment on Semantic Theory in the Cognitive Science of Conditional Reasoning," *Cognitive Science* 28(4):481–529.

Tamul D., Elson M., Ivory J., Hotter J., Lanier M., Wolf J., & Martínez-Carrillo N. 2020. „Moral Foundations' Methodological Foundations: A Systematic Analysis of Reliability in Research Using the Moral Foundations Questionnaire" (in press). Available at https://psyarxiv.com/shcgv/ (Last accessed on April 21, 2021).

Trups-Kalne I. & Dimdins G. 2017. "Relation between Social Conservatism, Moral Competence, Moral Orientations, and the Importance of Moral Foundations," *Mediterranean Journal of Social Sciences* 8(6):81–90.

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #4, pp. 59-72. DOI:10.14746/eip.2021.1.5

What, If Anything, Most Memorable Personal Moral Dilemmas Can Tell Us About Women's and Men's Moral Competence?

Marina A. Klimenko

(University of Florida, Gainesville, Florida; USA; mklimenko@ufl.edu)

Abstract: Most all popular moral reasoning or moral judgment tests are based on presenting subjects with two or more hypothetical moral dilemmas and asking them to make judgments on the moral quality of arguments supporting and questioning a protagonist's decision (e.g., the Moral Competence Test by G. Lind). Although these tests have been insightful by tapping some aspect of individuals' moral-cognitive schemas, moral maturity, or moral development, they also have limitations. Hypothetical moral dilemmas may be too abstract and impersonal, thus failing to create enough emotional salience. Learning more about real-life personally recalled moral dilemmas may reveal more about the individual's moral mind and experiences.
Objective. The current study was conducted to learn more about the personally experienced moral dilemmas, and how they relate to subjects' level of moral competence and gender.
Method. Subjects were asked to recall the most challenging personal moral dilemma; subjects completed the MCT test to measure moral competence.
Results. Among some of the findings was that for both, men and women, higher moral competence scores were positively correlated with recalling personal moral dilemmas where the choice had to be made between some altruistic (care for others) and selfish actions. For men, it was the risk of compromising one's status, whereas for women it was the risk of personal safety.

Keywords: moral competence; moral development; moral dilemmas of college students; sex-differences in moral judgement; MCT.

1. What, If Anything, Most Memorable Personal Moral Dilemmas Can Tell Us About Women's and Men's Moral Competence?

Most all commonly used standardized moral judgment tests are based on the cognitive developmental view of morality which assumes that moral thinking develops in stages and in parallel with one's cognitive maturation. Specifically, according to Lawrence

Kohlberg, who is arguably the father of moral psychology and of the original theory of moral development, morality progresses through three main and universal stages: pre-conventional, conventional, and post-conventional. To determine the structure of the individual's moral reasoning or the level of individual's moral maturity, hypothetical moral dilemmas are used for easy scoring and methodological standardization of assessment. As part of the procedure, participants are typically faced with two competing moral choices and forced to make one. In many of these standardized tests, including the original Moral Judgment Interview (MJI, Kohlberg 1984), each choice is accompanied by an argument justifying it. The subjects are scored on the choices that they make and the justification that they choose to determine subjects' moral orientation. A hypothetical moral dilemma that was developed by Kohlberg is whether a man should steal an expensive drug to save his wife's life (i.e., Kohlbergian Heinz dilemma). Kohlberg claimed that the choice and the justification for it, both, reflect the individual's moral cognitive structure. It may remain the same or it may continue to develop, for example with biological maturation or education.

The Moral Competence Test (MCT), which was designed by Georg Lind as a so-called Experimental Questionnaire (Lind 1982) and was influenced by the original MJI, incorporates two hypothetical moral dilemmas, mercy killing and worker's dilemma. The subjects are first asked to rate the protagonists' decisions for the choices that they would make in the dilemmas, and then asked to rate each decision for or against making a decision based on the reasoning that accompanies each choice. Thus, the key to the moral competence, according to Lind, is the ability to judge moral choices based on the quality of the arguments rather than on simply one's personal opinion for or against a particular decision.

While the ultimate goal of morality research is to be able to predict people's moral behaviour, even Kohlberg (1975) wrote that "one can reason in terms of principles and not live up to these principles." More recent studies suggest that reason may not be the only important factor that drives people's moral decisions. A dual-process model of moral decision-making, for instance, suggests that the decision is dependent upon two, working in parallel or competing with each other systems, cognitive (rational) and emotional (intuitive) (see Craigie 2011; Greene, Nystrom, Engel, Darley, & Cohen 2004). The conflict of a moral dilemma is presumably caused by the fact that personal moral values must be overridden in favour of a more rational option or vice versa. However, an argument can be made that a hypothetical moral dilemma may not always be able to elicit enough of a conflict if it is based on such abstract scenarios that both systems are not activated to some degree. Alternatively, a question can be raised as to how successful solutions of hypothetical moral dilemmas, as indexed by the higher score on a moral judgement test, is related to the subjects' personal experiences with real life moral dilemmas. Addressing these questions will provide important insights that can help tailor moral judgements

tests by choosing more nuanced moral dilemmas. Thus, the first goal of the current study was to examine the relationship between participants' moral competence (measured by MCT) and the recalled most memorable personal moral dilemmas reported by the same participants.

2. Personal vs. Impersonal/Hypothetical Moral Dilemmas

Studies with fMRI have found differences in the neural processing of hypothetical vs. real-life (personal) moral dilemmas. Personal dilemmas seem to be processed by the brain areas that are associated with emotion, and social cognition (e.g., medial frontal gyrus, posterior cingulate gyrus, and angular gyrus), suggesting that decisions of personal moral dilemmas are driven by socio-emotional factors; whereas hypothetical or impersonal moral dilemmas are processed by the cognitive areas of the brain (e.g., middle frontal gyrus, right and parietal lobe, bilateral), which points to higher-order processes and conscious deliberation during the decision of hypothetical moral dilemmas (Greene, Sommerville, Nystrom, Darley, & Cohen 2001). Furthermore, Greene *et al*. (2001) found that processing of impersonal moral dilemmas resembled the processing of non-moral dilemmas, for example, where one must choose between taking a train or a plane. Taken together, some hypothetical moral dilemmas may fail to evoke the emotional system of the brain for being less emotionally salient. There are potentially other differences between personal and impersonal moral dilemmas. These differences were examined in the present study as part of the second goal.

3. Difficult vs. Easy Moral Dilemmas

Typically, a moral dilemma is a situation where one must decide between at least two competing moral values. Such dilemmas are challenging because moral values cannot always be objectively placed in the order of superiority (Trainer 1982). Greene and colleagues (2004) argued that some of the tensions come from needing to suppress a potent negative emotional response, when one knows he/she may potentially violate one's personal value(s), in order to make a decision that will benefit more people in the long run, or vice versa. These two different approaches to moral dilemma solutions are known as utilitarian, when an individual chooses to maximize benefits and minimizes the costs, and deontological, choosing to follow one's moral intuition. Greene *et a*l. (2004) further argued that more difficult moral dilemmas are the ones where the emotional and the cognitive decisions are at greater odds. An example that was used in their study was the so-called crying baby dilemma, where the participants had to decide whether to smother one's own baby who is crying loudly and will attract enemy soldiers. If one does, that person will save herself/himself and the others, but if one chooses not to smother her own baby, everyone will die, including the baby.

The study found that, unlike deciding on a simpler moral dilemma, a decision on the difficult (crying baby) dilemma was associated with activity bilaterally in both the anterior dorsolateral prefrontal cortex (DLPFC) and inferior parietal lobes, which are known for cognitive processing. There was also more activity in the anterior cingulate cortex ACC, known for processing conflict (Greene *et al.* 2004). Greene *et al.* (2004) referred to such emotionally salient dilemmas as personal, however, one can still argue that they are still hypothetical as subjects may not personally relate to the presented situations. Furthermore, emotional salience may still differ across the subjects when they are presented with the crying baby dilemma. It remains unknown what individuals' enduring memories of certain personally experienced dilemmas may reveal about their moral maturity.

From the non-neuroscience perspective, the degree of difficulty or the tension between the competing moral values can be evaluated by considering the motivations behind the moral decisions (e.g., Lewis & Mitchell 2014). Specifically, two pitted moral values can be both driven by altruistic motivations, as in the crying baby dilemma, where two conflicting values can be both driven by care (care for well-being of one's baby or care for well-being of innocent people). A moral dilemma where an individual is deciding on whether to cheat on taxes (to save money) or not to cheat (to avoid getting caught) is driven by two selfish motivations. Finally, a moral dilemma may entail an altruistic and a selfish motivation, as in contemplating to donate money (to help another person) or not (to keep money for yourself). These categorizations of moral dilemmas allow placing them in the order of moral superiority, from most morally conflicting (i.e., difficult to solve) to the least morally conflicting (easy to solve), where a dilemma entailing two altruistic motivations reflects the highest level (most difficult to resolve for a highly moral individual) and a moral dilemma driven by two selfish motivations to be the lowest (the easiest to resolve for a highly moral individual). The second question that the present study addressed was what type of moral dilemmas would predict higher moral competence. It was hypothesized that subjects who would recall a moral dilemma involving two altruistic moral motivations as most memorable (i.e., most difficult based on the motivation categorization) or involving an altruistic and selfish motivations (i.e., high emotional salience) would have the highest level of moral competence; whereas the subjects, who would report a moral dilemma entailing two selfish motivations to be most memorable, would have the lowest moral competence.

4. Sex Differences in Reported Moral Dilemmas and Moral Competence

On one hand, there seems to be no strong empirical evidence from past research using various standardized moral reasoning tests that sex-related differences exist in moral reasoning. For example, a meta-analysis by Walker (1984) found no significant differences between young children, adolescents, and adults in moral reasoning attributed

to sex. In a longitudinal study, Walker (1989) examined moral reasoning of participants, ranging in age between 5 and 63, over a two-year period. All participants discussed hypothetical moral dilemmas and a personal moral dilemma. The study found virtually no significant differences between female and male subjects in their moral orientation or reasoning using hypothetical moral dilemmas. However, differences in moral reasoning did emerge when personal moral dilemmas were evaluated. In other words, significant differences in moral reasoning were only attributed to the types of moral dilemmas that were used for assessment.

In a more recent study, Capraro and Sippel (2017) examined sex differences in judging three different moral dilemmas, one where utilitarian and deontological moralities are at the greatest odds, due to greater emotional salience and greater violation of practical imperative, a concept coined by Immanuel Kant who defined morality as " … act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means'' (Kant 1785, 4:429). This type of moral dilemmas was categorized in the study as personal. The second moral dilemma was referred to as impersonal, because it didn't introduce to the same degree a moral conflict between utilitarian and deontological courses of action. The third dilemma was categorized as intermediate. It had a conflict that violated practical imperative as the personal dilemma but was emotionally less salient. The results showed that women were more likely to take deontological moral approach when judging the personal moral dilemma; no differences were found in judging the intermediate moral dilemma, suggesting that the sex differences were driven by the emotional salience rather than by the violation of practical imperative. Therefore, it is possible that differences between female and male respondents will emerge if personal, presumably emotionally more salient, moral dilemmas are probed.

5. Methods

5.1 Sample

A total of 339 participants were recruited for the study in exchange for extra credit from four Psychology classes, which were all taught by the author. The majority were first- and second-year students; 274 were female students. In class #1, students completed Moral Competence Test (MCT) (Lind 2014) at the beginning and the end of the semester; in three other classes students completed the MCT test at the end of the semester only. For all subsequent analyses, only the scores on the MCT test, which was administered at the end of the semester, were used for all participants. In addition, at the end of the semester, all participants were asked to recall one most memorable (i.e., challenging) moral dilemma that they had to face. There was a slight difference in the prompt that was

given to participants in class #1 and the rest of the participants. Specifically, the students in the former were not explicitly told that they could report an impersonal moral dilemma if they could not recall a personal moral dilemma. Despite the difference in the prompts, some respondents reported a hypothetical or an impersonal moral dilemma (a moral dilemma that somebody they knew had to face) in all four classes.

5.2 Coding scheme

All moral dilemmas were coded based on

(1) Whether they were personal or not personal,

(2) The time of its occurrence (i.e., before college or during college/ongoing), and

(3) The competing moral motivations that created the conflict.

The latter was based on the coding scheme by Lewis and Mitchell (2014), which categorized moral conflicts in terms of motivational drives behind choosing one or another act. These motivations were categorized as altruistic or egoistic motivations, each of 5 subcategories (see Lewis & Mitchell 2014 for details of the coding scheme). Specifically, the five possible altruistic motivations were: care, fairness, in-group loyalty, authority, and purity. The five possible egoistic motivations were: immediate physiological needs, safety, affiliation, status, and mating.

A second coder independently coded 15 % of all moral dilemmas. Cohen's kappa of inter-rater reliability was .423 for altruistic motivations and .494 for egoistic motivations, both considered fair in terms of reliability. Cohen's kappa for personal/impersonal codes was 1.00, which is a perfect agreement.

Three broad types of moral motivations were created based on the two competing motivations. These could be of three types:

(1) Moral dilemmas with two competing altruistic motivations (e.g., to help a group of unknown people or to help a best friend);

(2) Moral dilemmas with one altruistic and one egoistic competing motivations (e.g., to avoid personal danger or to help a friend), and

(3) Moral dilemmas with two competing egoistic motivations (e.g., to report a stolen wallet to avoid being judged or to keep the wallet and spend the money) (see Lewis & Mitchell 2014 for more details).

For the follow-up analyses, 25 moral dilemma combinations were created as separate variables (dummy coded) to examine each dilemma combination associations with the subjects' sex and moral competence scores.

6. Results

There were no differences in the number of recalled personal or impersonal moral dilemmas between the four classes (X^2 (3, N = 335) = 5.17, p = .160. However, more students (N = 222) recalled a personal moral dilemma, as prompted by the instructions,

X^2 (1, N = 335) = 35.4, p = .000. Significantly more dilemmas were recalled from the time of being in high school than from either earlier or more recent times (X^2 (4, N = 177) = 97.0, p = .000).

6.1 Reported moral dilemmas

The majority of moral dilemmas where the decision had to be made between choosing an altruistic and an egoistic act (37%, N = 186), or two altruistic acts (25%, N = 127); the least commonly reported dilemmas involved two egoistic actions (5%, N = 25). These differences were statistically significant, X^2 (2, N = 338) = 117.0, p = .000.

As far as what motivational domains were overrepresented in the moral dilemmas, a goodness of fit test indicated significant differences between the domains, X^2 (9, N = 678) = 794, p = .000. The residuals suggest that concerns of care were the most overrepresented (i.e., the most reported) altruistic motivations, while concerns for affiliation were the most overrepresented (i.e., the most reported) egoistic motivations. Purity and mating were the most underrepresented (i.e., the least reported) altruistic and egoistic motivations (see Table 1 for details).

Motivations	Observed N	Expected N	Residual
Care	264 (39%)	67.8	196.2
Fairness	28 (4%)	67.8	-39.8
Ingroup	34 (5%)	67.8	-33.8
Authority	112 (16%)	67.8	44.2
Purity	5 (0.7%)	67.8	-62.8
physiological needs	59 (9%)	67.8	-8.8
Safety	28 (4%)	67.8	-39.8
Affiliation	95 (14%)	67.8	27.2
Status	48 (7%)	67.8	-19.8
Mating	5 (0.7%)	67.8	-62.8
Total	678 (100%)		

Table 1: Individual Motivational Domains.

6.2 Personal vs. hypothetical/impersonal moral dilemmas

To examine the differences between personal and impersonal moral dilemmas, a chi-square test was computed. This test revealed a significant difference, X^2 (2, N = 334) = 24.9, p = .000. Specifically, more personal moral dilemmas (61%) were about choosing between an altruistic and an egoistic act; whereas more impersonal dilemmas (55%) were about choosing two altruistic actions.

6.3 Sex differences and the link between personally recalled moral dilemmas and moral competence

Next, potential sex differences in the reported dilemmas and the level of moral competence were explored. First, an independent samples t-test was computed with sex and moral competence scores. A significant difference was found, with male students having a higher moral competence level (M=17.9, SD=11.8) than the female students

(*M*=13.9, *SD*=10.3), *t*(385) = 2.1, *p* = .00. To examine potential differences in the reported types of moral dilemmas between female and male students, a chi-square test was computed. The test revealed that about 55% of female students reported moral dilemmas that involved an altruistic and an egoistic act, whereas 53% of male students reported moral dilemmas that involved two altruistic acts. These differences were approaching a statistical significance, X^2 (2, *N* = 323) = 5.25, *p* = .07. However, for both sexes, moral dilemmas that involved two egoistic choices were rare.

Given that there were clear sex-related differences in the reported moral dilemmas and the moral competence measures, a two-way ANOVA was computed first with sex and the types of moral dilemmas as fixed factors and moral competence scores as the dependent variable. The results showed that neither the types of moral dilemmas (*F* (2, 296) = 1.99, *p* = .138) nor the sex of the respondents (*F* (1, 296) = .533, *p* = .466) were significant predictors of moral competence. When the test was rerun without the sex as a one-way ANOVA, the model was statistically significant (*F*(2, 310) = 4.64, *p* = .010). A Tukey post hoc test revealed that moral competence of those reported a moral dilemma with two egoistic motivations was statistically significantly lower (*M*=8.51) than those who reported a moral dilemma that entails an egoistic and an altruistic motivation (*M*=15.6, *p* = .008) or two altruistic motivations (*M*=14.1, *p* = .053). No other statistically significant differences were found.

Next, to find out if sex and specific kinds of moral motivations would predict the level of moral competence, a series of two-way ANOVAs were computed with each possible combination of altruistic-altruistic and altruistic-egoistic motivations (since these types of moral dilemma were associated with higher moral competence, and they were the most frequently reported dilemmas). In each model, sex and each moral dilemma combination were entered as fixed factors, and moral competence as the dependent variable. For parsimony, care domain was combined with fairness domain as the coders viewed them very similar. Purity and mating domains were not included in the analyses since they were very infrequent. Moral dilemma types reported by less than 20 people were not considered for these analyses either (see in Table 2).

Moral Dilemma Combinations	Total Number of Reported
Altruistic-Altruistic Motivations	
Care-Care (choosing between caring for the well-being of one or more people)	57
Care and In-group (choosing between caring for well-being of a person vs. staying loyal to one's group)	6
Care-Authority (choosing between caring for well-being of a person vs. choosing the authority/rules/laws)	41

In-Group and In-Group (choosing between staying loyal to one or another group)	4
In-Group and Authority (choosing between staying loyal to one's group or following the authority/rules/laws)	8
Authority and Authority (choosing between two different authority/ rules/laws)	0
Altruistic-Egoistic Motivations	
Care-Physiological needs (choosing between caring for well-being of a person or personal physiological needs)	36
Care-Safety (choosing between caring for well-being of a person or personal safety)	22
Care-Affiliation (choosing between caring for well-being of a person or personal belonging/being affiliated with a group/friendship)	35
Care-Status (choosing between caring for well-being of a person or personal status)	22
In-group and Physiological needs (choosing between staying loyal to one's group or personal physiological need)	1
In-group and Safety (choosing between staying loyal to one's group or personal safety)	1
In-group and Affiliation (choosing between staying loyal to one's group or personal affiliation)	6
In-group and Status (choosing between staying loyal to one's group or personal status)	2
Authority-Physiological needs (choosing between following authority/ rules/laws or personal physiological needs)	10
Authority-Safety (choosing between following authority/rules/laws or personal safety)	3
Authority-Affiliation (choosing between following authority/rules/laws or personal affiliation)	27
Authority and Status (choosing between following authority/rules/laws or personal status)	13

Table 2: Number of moral dilemma types reported.

Only two models were found significant in predicting moral competence.

Specifically, in the model with the dilemmas that involved choosing between care (an altruistic motivation) and personal status (an egoistic motivation) there was a statistically significant effect of sex (F (1, 299) = 5.38, p = .02), and the interaction between the sex and the care/status moral dilemma interaction was approaching statistical significance, F (1, 299) = 2.61, p = .10. Specifically, male students who reported this type of moral dilemma had a higher moral competence (M=24.3, SD=9.4) then who reported having a different type of moral dilemma (M=16.8, SD=10.3) and higher than the female who reported having the same (M=11.2, SD=8.1) or a different type of a moral dilemma, compared to their female counterparts (M=14.5, SD=10.5). However, these findings must

be interpreted with caution, as only 3 out of 46 male respondents responded having this type of dilemma.

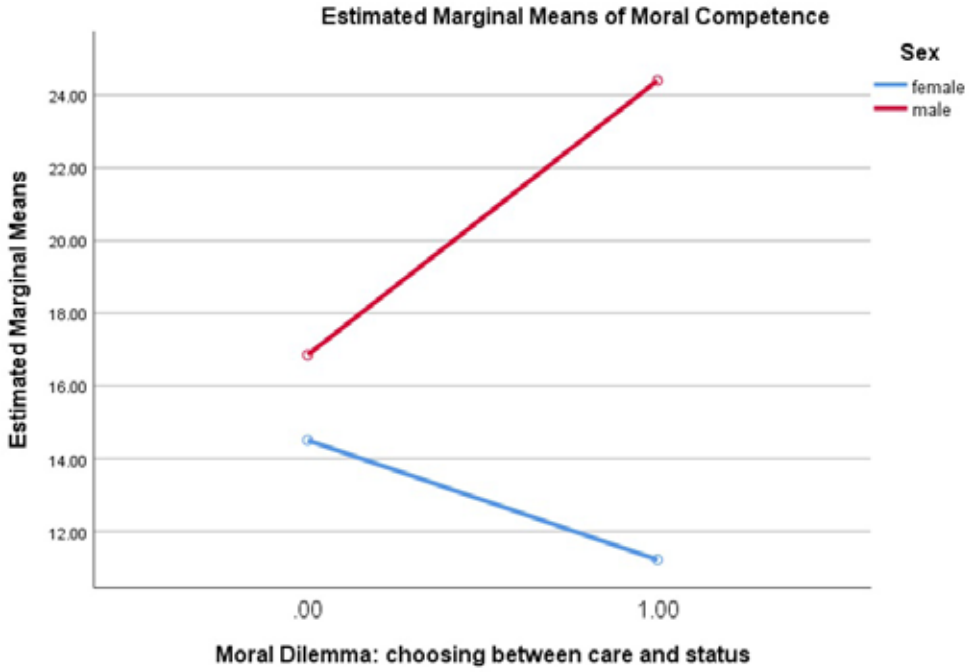


Figure 1: Interaction between sex and care/status moral dilemma.

For a model with care (an altruistic) vs. personal safety (an egoistic), both, reporting this type of moral dilemma (F (1, 300) = 9.08, p = .00), and sex, F (1, 300) = 4.79, p = .02, were significant predictors of moral competence. Specifically, students who reported this type of moral dilemma had a higher moral competence (M=21.1, SD=14.1). Furthermore, they were all female respondents.

7. Ad hoc Analysis of Sex Differences Between Reported Personal and Hypothetical Moral Dilemmas

In light of the findings that more men than women reported more moral dilemmas that involved two altruistic decisions, two additional chi-square tests were performed with two specific types of moral dilemmas and sex as the predictors: care vs. care and care vs. authority. These were, by far the most frequently reported dilemma in the altruistic/altruistic category. The test revealed a significant sex difference but only for care vs. care moral dilemmas, X^2 (1, N = 275) = 4.80, p = .03. Specifically, 28% of male subjects reported this type moral dilemma as compared to 15% of the female subjects. Next, another chi-square test was performed, between care/care moral dilemma variable and personal/hypothetical category variable (both dummy coded variables). The results revealed a significant difference, X^2 (1, N = 335) = 26.3, p = .00. Specifically, 63% of care/care dilemmas were hypothetical.

8. Discussion

A moral dilemma is only a dilemma if the one debating it perceives it to have a conflict between two or more values. For example, choosing between donating money to a charity or spending it on yourself is only a moral dilemma if one is contemplating both actions; it is not a dilemma if a person is not even considering donating money or if a person feels strongly about donating money that he/she doesn't feel it to be a hard decision. In any of these cases, it reveals something about a person's moral character. Typical standardized moral judgement tests use hypothetical moral dilemmas to assess people's level of moral maturity. Whether and how moral standardized scores are actually related to people's personal moral dilemmas had not been researched enough or at all, to my knowledge. Thus, the current study was the first of its kind to examine the potential connection between the measure of moral competence (MCT, Lind 1982; 2014) and people's perception of a difficult personal moral dilemma that they had to face with. First, the study found that, at least for college students, most challenging moral dilemmas were the ones where they had to choose between either two different morally competing but altruistic actions or between choosing an action driven by a personal gain or an altruistic motivation. Very few college students reported having difficulty with deciding between two selfishly motivated actions; and thus, few reported such moral dilemmas. This confirms prior evidence that education fosters higher level of moral development. It is also possible that more individuals with higher level of moral maturity choose to and/or get accepted to higher education. Finally, this may be a function of age – i.e., young adults may all tend to consider moral dilemmas as those that require to decide between two or more altruistic actions or between an altruistic and a selfish one.

Second, more college students recalled moral dilemmas that occurred in high-school rather than more recently. This highlights the emotional salience of the adolescent stage and the issues that may arise during this time of development. Another notable finding was that *difficult* personal dilemmas were mainly about choosing between self-interests and altruism. Relinquishing personal interests must be emotionally salient and therefore, creates a memorable moral dilemma for many individuals. On the other hand, *difficult* hypothetical or impersonal moral dilemmas were mainly about choosing between two altruistic actions. These types of moral dilemmas maybe emotionally salient because they are designed to have no correct solution, and thus, virtually unsolvable. In any case, both types of moral dilemmas were positively correlated with higher moral competence.

The results of the study also revealed some interesting sex differences. Specifically, female subjects reported more personal rather than impersonal moral dilemmas whereas male subjects were more likely to report an impersonal or a hypothetical moral dilemma. There are two potential explanations for this. First, this may be related to known sex differences in self-disclosure, defined as "intentional and voluntary verbal utterance that conveys personal information to another within a specific social context" (Papini, Farmer,

Clark, Micka, & Barnett 1990). Research on self-disclosure suggests that females tend to share their feelings and to self-disclose more than their male counterparts in person and, possibly, on social media (e.g., Bond 2009; Papini *et al.* 1990). Therefore, female subjects may feel more comfortable reporting sensitive information such as personally experienced moral dilemma than men. A second explanation may be related to what Carol Gilligan theorized to be sex-related differences in the modes of thinking and feeling about moral and other social issues, which stem from the differences in social experiences between men and women (e.g., Gilligan 1986). As Gilligan claimed, men tend to think of morality in terms of justice orientation, which more likely to take on a depersonalized approach to moral reasoning; whereas women reason through the lenses of relationships and care for others.

Hypothetical moral dilemmas may evoke justice oriented moral mode of thinking which male respondents find more fitting their moral mode of thinking; whereas personal dilemmas maybe more grounded in viewing morality through the self, others and interpersonal relationships and thus may be more memorable to female subjects. Alternatively, men with higher moral competence may find hypothetical moral dilemmas particularly challenging and memorable because they tend to have no good rational solution. For example, a study by Fumagalli *et al*. (2010) found that more men gave utilitarian (rational) solutions to emotionally more salient moral dilemmas than women, suggesting that men tend to take a more rational approach to solution of a moral dilemma.

Still, another important finding of the current study was that both, men and women, who reported a personal moral dilemma that required to choose between personal self-interests (i.e., egoistic motivation) and altruistic action of care, had the highest level of moral competence than the rest of the subjects. Where they differed was in what they were willing to or could potentially compromise when choosing the act of care. For men, it was the risk of losing one's status (e.g., not to win a scholarship or not to go to a prestigious university), whereas for women it was the risk of personal safety (e.g., giving a ride to a stranger).

These findings are insightful for a couple of reasons. First, the similarity between men and women in that, for both sexes, care for others was one of the most frequent moral motivations that created a memorable moral dilemma is noteworthy and supports Kohlberg's view that women and men are more similar than different in their mode of moral thinking. It also supports the notion that that caring for human life is one of the rudimentary human moral values (e.g., Haidt & Joseph 2007). This finding also adds more insight to the findings of the study by Capraro and Sippel (2017), where the key difference in the moral judgment between men and women was the emotional salience, and not the Kantian practical imperative, of the personal moral dilemma. The current study suggests that emotional salience may be equally important for men and women; it may just be evoked by different moral dilemmas. For men, it appears to involve risking one's status, whereas for women, this maybe the risk of losing personal safety. Therefore, a conclusion

can be made that emotional engagement is the key to promoting higher moral reasoning and judgement (e.g., Greene *et al*. 2001).

The study is not without limitations. The main one is the homogeneity of the sample. The subjects were mainly college age students, at a major US university. Furthermore, the majority of the participants were females. More follow up research is necessary with older and more diverse subjects (different educational and cultural backgrounds) to be able to generalize the results of the findings to broader population.

References

Bond B. J. 2009. "He Posted, She Posted: Gender Differences in Self-disclosure on Social Network Sites," *Rocky Mountain Communication Review* 6(2):29–37.

Capraro V. & Sippel J. 2017. "Gender Differences in Moral Judgment and the Evaluation of Gender-specified Moral Agents," *Cognitive Process* 9. DOI: 10.1007/s10339-017-0822-9.

Craigie J. 2011. "Thinking and Feeling: Moral Deliberation in a Dual-process Framework," *Philosophical Psychology* 24:53–71. DOI: 10.1080/09515089.2010.533262

Furmagalli M., Ferrucci R., Mameli F., Marceglia S., Mrakic-Sposta S., *et al.* 2010. "Gender-related Differences," *Cognitive Process* 11:219–226.

Gilligan C. 1986. "Reply by Carol Gilligan," *Signs* 11:324–333.

Greene J. D., Sommerville R. B., Nystrom L. E., Darley J. M., & Cohen J. D. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293:2105–2108.

Greene J. D., Nystrom L. E., Engel C. L., Darley J. M., & Cohen J. D. 2004. „The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44:389–400.

Haidt J. & Joseph C. 2007. "The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-specific Virtues, and Perhaps Even Modules," in P. Carruthers, S. Laurence, & S. Stich (Eds.), *Evolution and Cognition. The Innate Mind. Vol. 3: Foundations and the Future* (p. 367–391). Oxford University Press. URL: https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

Krebs D. L., Denton K., & Wark G. 2006. "The Forms and Functions of Real-Life Moral Decision-Making," *Journal of Moral Education* 26:131–145.

Lewis R. M. & Mitchell N. 2014. "Egoism Versus Altruism in Television Content for Young Audiences," *Mass Communication and Society* 17:597–613.

Lind G. 1982. "Experimental Questionnaires: A New Approach to Personality Research," in A. Kossakowski & K. Obuchowski (Eds.), *Progress in Psychology of Personality* (pp. 132–144). Amsterdam, NL: North-Holland.

Lind G. 2014 (March 24). "Moral Competence Test (MCT), URL: moralcompetence.net

Lind G. 2016. *How to Teach Morality. Promoting Deliberation and Discussion, Reducing Violence and Deceit*. Berlin: Logos.

Papini D. R., Farmer F. F., Clark S. M., Micka J. C., & Barnett J. K. 1990. "Early Adolescent Age and Gender Differences in Patterns of Emotional Self-disclosure to Parents and Friends," *Adolescence* 100:959–976.

Rest J. R. 1979. *Development in Judging Moral Issues*. Minneapolis: University of Minnesota Press.

Singer N., Kreuzpointner L., Sommer M., Wüst S., & Kudielka B. M. 2019. „Decision-making in Everyday Moral Conflict Situations: Development and Validation of a New Measure," *PLoS ONE* 14(4). Article #e0214747. DOI: 10.1371/journal.pone.0214747.

Trainer F. E. 1982. *Dimensions of Moral Thought*. New South Wales: NSW University Press Ltd.

Trainer T. 1991. *The Nature of Morality*. Avebury, England: Academic Publishing Group.

Walker L. 1984. "Sex Differences in the Development of Moral Reasoning: A Critical Review," *Child Development* 55:677–691.

Walker L. 1989. "A Longitudinal Study of Moral Reasoning," *Child Development* 60:157–166.

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #5, pp. 73-83. DOI:10.14746/eip.2021.1.6

The Effect of Moral Competence on Online Conformity Behavior

Aida Mofakhami

(Allameh Tabataba'I University, Teheran; Iran; aida.mofakhami@gmail.com)
ORCID: 0000-0002-3390-6488

Abstract: Moral Competence, defined as the ability to solve conflicts on the basis of shared moral principles through cooperation rather than through violence, deceit and power, has received little attention among different psychological approaches; despite its importance in predicting many of our social interactions. The purpose of this study was to investigate the effect of moral competence on online conformity behavior. 217 students from universities in Teheran were selected for a quasi-experimental study. First, participants' moral competence was measured with the online Moral Competence Test (MCT) by G. Lind (1978/2019). Then the subjects participated in an online version of an Asch type experiment in which conformity was induced. The results showed a clear conformity behavior in the use of the internet. An average of 32.09% of participants conformed to each critical question. When compared to Asch`s line judgment task, the mean conformity in this experiment was lower, but still significant enough to indicate conformity behavior (36.8% compared with 7.4%), which might stem from the online situation, in which some other variables like the deindividuation effect might influence this difference. The results also indicated that there was a weak but negative correlation between moral competence and conformity behavior. The results confirm our hypothesis weakly; subjects with higher moral competence tended to show lower conformity. If the results could be replicated, it would imply that conformity is not a general and stable trait of people, as Asch assumed, but depends on people's level of moral competence, which can be fostered through education.

Keywords: morality; moral competence; social conformity; online conformity; MCT; experimental research with university students.

Introduction

Conformity as a psychological term is defined as an influence resulting from one's willingness to accept others' opinions about reality (Asch 1955) . Although it is human nature to follow others, conformity can lead to very dangerous behaviors. As Zimbardo (2007) has noted, conformity is a strong group psychological mechanism that can make

people behave inhumanly. On the other hand, conformity as a social mechanism can also serve as a prosocial behavior, helping group dynamics and more effective communication (Bond & Smith 1996). The examples of this behavior occur in almost all our everyday life contexts, e.g., when we stand while our national anthem is played as others do so, or when we "like" others' posts on social media such as Instagram just because they have been liked by a great number of other people; and generally, when we confirm to ideas just because the majority has accepted them, not because we truly believe in them. In such situations we are simply conforming. As the internet has penetrated every facet of our lives and has become inseparable from day-to-day conduct, it is necessary to take into account this emerging context when studying such variables in the social interactions of individuals. With this in mind, the current study was designed to take an online form, in order to investigate online conformity behavior.

Solomon Asch (1958) was one of the first psychologists to study conformity. He used a lab experiment in which a group of eight persons participated in a "line judgment" task, when in fact only one of them was the real participant, and the other seven were confederates/stooges who acted as participants, while the real participant was unaware of this. During the experiment, each student viewed a card showing a target line on one side and three comparison lines on the other side. Each person in the room had to choose aloud which comparison line (A, B or C) was similar to the target line. The answer was always obvious. The real participant was always the last person who had to give their answer, and the confederates had agreed in advance to give an obviously wrong answer in most trials (critical trials). On average, about one third of the participants who were placed in this situation sided with the clearly wrong majority in the critical trials. Asch's experiment (1955, 1956, 1958) also featured a control condition where there were no confederates, only "real participants". In Asch's line judgment task, the test subjects complied on average with the majority's wrong judgment in 36% of their selections (Rosander & Eriksson 2012). Although conformity as a face-to-face behavior has been widely studied by many previous researchers (Deutsch & Gerard 1955; Bond & Smith 1996; Baumeister 1982; Janes & Olson 2000; Goeree & Yarive 2015), to this day there are not many studies that have investigated conformity behavior in non face-to-face situations.

With regard to recent studies on conformity in CMC (computer mediated communication) conditions, Rosander and Eriksson (2012) authored *Conformity on the Internet* and *The Role of Task Difficulty and Gender Differences*. These are well-designed studies, from which the present study draws inspiration. Rosander and Eriksson used a web-based survey as the social context for their study, and the results showed that 52.6% conformed at least once, with an average 13.0% of participants conforming on each critical question. The conformity increased with higher task difficulty, and no difference was found between men and women in terms of their conformity behavior. The study discussed some reasons for this form of conformity behavior, based on theories and previous studies,

such as turning to the group for guidance, avoiding social isolation and protecting one's self-esteem (Rosander & Eriksson 2012). However, Rosander and Eriksson placed more emphasis on environmental factors that influenced conformity, such as social isolation. The importance of more internal variables, which seem to differ due to education, have received little attention so far, yet they may also play an essential role in conformity behavior. In other words, conformity may not be as static as Asch assumed.

In the present study, Moral Competence comes into play as another variable. We assume it has an impact on conformity behavior, on the basis of its theoretical underpinnings and its definition. This concept was first introduced by Georg Lind (1978), the German psychologist and philosopher. He defines moral competence "*as the ability to solve conflicts and problems on the basis of shared moral principles through thinking and discussion rather than through violence, deceit and power*." Specifically, it is the ability to rate the arguments of others with regard to their moral quality rather their opinion-agreement. This study considers moral competence as a key variable which plays an essential role in the extent of an individual's conformity behavior.

Conformity may also play a role when people behave immorally when others ask them to do so, even when their own moral principles would not allow this. Stanley Milgram (1963) has shown that two thirds of his participants gave other subjects electroshocks in an alleged learning experiment, even though they thought this was wrong, as they said afterwards. Lawrence Kohlberg (1984) repeated Milgram's experiment, showing that participants with high moral competence refused to obey this immoral instruction. They seem to be immune to conforming with immoral orders. Perhaps the influence of moral competence on conforming behavior can also be demonstrated with the current experiment.

1. Main Hypotheses

As was mentioned above, there is evidence for conformity behavior in CMC, but the strength of such behavior is considerably lower when compared with face-to-face situations, as investigated by Asch (Rosander & Errikson 2012; Bond & Smith 1996). Therefore, we replicated this method in order to investigate whether the previous results are consistent in different cultural settings. For this reason, only Iranian students participated in this experiment.

On the other hand, as the definition of moral competence suggests, the more people were able to judge arguments on the basis of their own principles (and not under the influence of other opinions), the less they will conform to the wrong majority options. In other words, people with higher moral competence are less likely to express conformity behavior in conditions characterized by social pressure. Therefore, we propose:

(1) Conformity behavior would occur in an online context;

(2) People with high moral competence are less likely to succumb to the pressure to conform.

2. Research Methodology

2.1 Participants

There was a total of 217 participants, consisting of 140 women and 77 men (71.7% females and 28.3% males), ranging from 18 to 36 years old, and they all participated voluntarily. They mostly came from the website to which this experimental design was uploaded (www.Ravanhami.ir) and received a cash credit as a reward for their participation at the end of the experiment. Furthermore, regarding ethical considerations, all the participants received an email explaining the true aim of the research after the deadline for the survey.

The population of this study compromised undergraduate students of three universities in Teheran in the Winter Semester of 2017-2018.

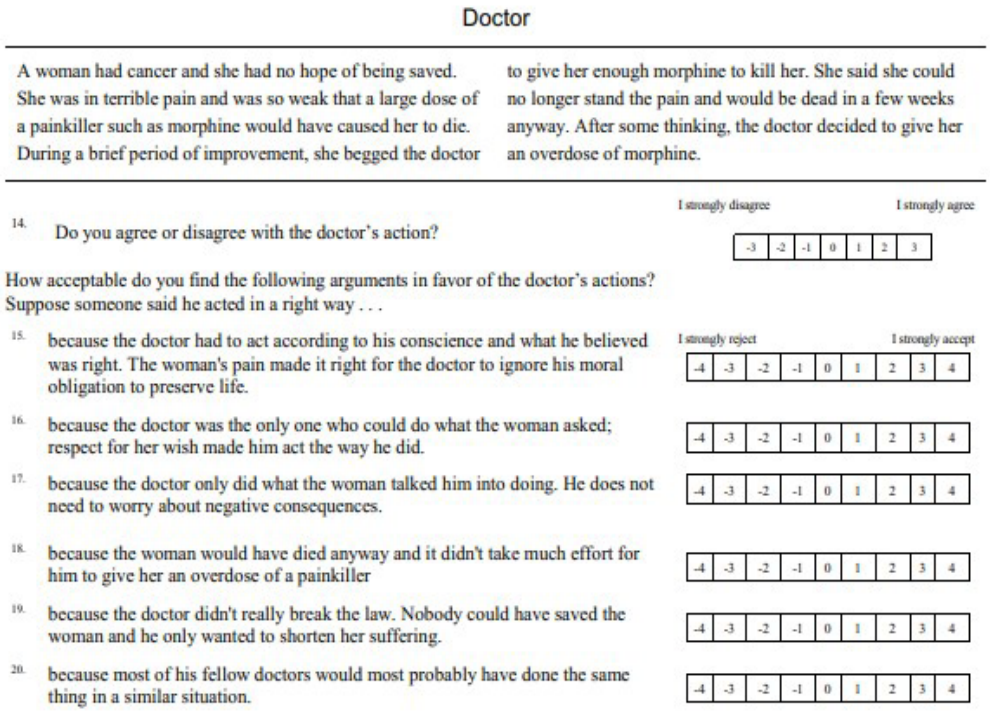
The subjects were divided into two separate groups (one control group and one conformity group). 10 participants were eliminated from the sample because they were under 18 or did not complete the whole experiment.

2.2 Materials

2.2.1 MCT (Online form)

The Moral Competence Test (MCT) contains 24 arguments pro and contra the protagonists' decision in two dilemma situations which are to be rated with regard to their acceptability. They are first asked to express their opinion about the actor's decision (the first six arguments are in support of the actor's decision and the other 6 arguments contradict his decision) and then, regardless of their opinion, they are asked to rank the arguments on a Likert scale ranging from -4 (strongly disagree) to +4 (strongly agree). Each argument was designed to represent one of Kohlberg's six type of moral orientation (Lind2019). Depending on the pattern of their answers, the respondents' moral competence is scored and receives a score ranging from 0 (no moral competence at all) to 100 (very high moral competence). The MCT is suitable for people over 10 years old.

The implicit task of the MCT is to rate the arguments with regard to their moral quality instead of in terms of whether participants agree with the arguments. That is, this task was not made explicit to the subjects. In this study, the Persian translation of the MCT by Saeidi (2011) was used in an online form, which was available on a web site named www.ravanhami.ir.



Doctor

A woman had cancer and she had no hope of being saved. She was in terrible pain and was so weak that a large dose of a painkiller such as morphine would have caused her to die. During a brief period of improvement, she begged the doctor to give her enough morphine to kill her. She said she could no longer stand the pain and would be dead in a few weeks anyway. After some thinking, the doctor decided to give her an overdose of morphine.

14. Do you agree or disagree with the doctor's action?

I strongly disagree | -3 -2 -1 0 1 2 3 | I strongly agree

How acceptable do you find the following arguments in favor of the doctor's actions? Suppose someone said he acted in a right way . . .

I strongly reject | -4 -3 -2 -1 0 1 2 3 4 | I strongly accept

15. because the doctor had to act according to his conscience and what he believed was right. The woman's pain made it right for the doctor to ignore his moral obligation to preserve life.

16. because the doctor was the only one who could do what the woman asked; respect for her wish made him act the way he did.

17. because the doctor only did what the woman talked him into doing. He does not need to worry about negative consequences.

18. because the woman would have died anyway and it didn't take much effort for him to give her an overdose of a painkiller

19. because the doctor didn't really break the law. Nobody could have saved the woman and he only wanted to shorten her suffering.

20. because most of his fellow doctors would most probably have done the same thing in a similar situation.

Figure 1: MCT by G. Lind (1978/2019). Doctor-Dilemma with six supporting arguments as an example.

2.2.2 Conformity situation (Web-based survey)

In order to induce conformity behavior, the author designed an experimental situation which was methodologically a replication of Rosander and Eriksson (2012) but customized the question in a way that matches to the participants' cultural background. In this situation, the participants were randomly assigned to one of the two groups, in which they were asked to answer 40 questions about general knowledge in four areas including, chemistry, history, literature and geography. Participants in the experimental group were presented with a fake diagram showing that the great majority of the previous participants had chosen an obviously wrong option, while the control group received no diagrams for the same questions (Figure 2). Conforming answers were given the value of one and nonconforming answers were given the value of zero.

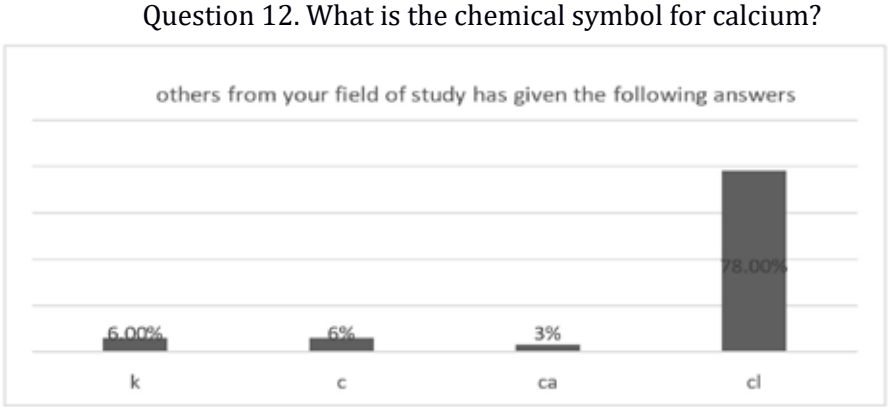


Figure 2: an example question shown to THE experimental group (critical question).

a. k
b. C (conformity answer)
c. Ca (correct answer)
d. Cl

2.3 Research design and procedure

The study was quasi-experimental, with an experimental condition. In fact, the fake diagrams played the role of confederates in Asch's classic experiment, because they showed that the majority of participants had selected the obviously wrong answer.

In about 40% of the cases, the correct description of the answer's distribution was displayed on diagrams called "neutral question" and the distribution in the other 60% of the diagrams was fabricated, showing the majority chose an incorrect answer, called "critical questions" (Rosander & Eriksson 2012). The reason for having 40% of neutral question was to avoid raising too much suspicion in the participants, since a high level of suspicion seems to have a negative influence on conformity behavior and results in methodological problems (Stang 1976).

The procedure of this study included 3 steps:

Step 1: All the participants completed one page of demographic questions (age, gender and education level).

Step 2: They answered the online form of MCT questions before starting the experimental phase.

Step 3: They were assigned to one of the two groups and were asked to answer 40 items in the general knowledge test. The order of questions was identical in both the control and experimental groups.

In order to prevent the web search effect for the experimental part of the research, the participants had only 30 seconds to answer each question and they could only go forward during the experiment. They were not allowed to use the "back" option during the test.

3. Research Results

For the statistical analysis of the two main hypotheses of this study, PSPP were used. No significant differences were found regarding the age, gender differences and educational levels between the conformity and control groups.

3.1 Conformity behavior would occur in an online context

With regard to this assumption, we expected a significant difference in conformity levels between the control and experimental groups. For the twenty-two critical questions, the mean number of answers in accordance with the manipulation in the conformity group (M=7.482, SD=4.421, N=112) was higher than the control group (M=3.284, SD=1.766,

n=96). As can be seen, the difference between the mean scores of the experimental and control groups suggests that in this experimental situation conformity does occur; that is, even people who are alone when using the internet still succumb to the pressure of the majority and conform (Figure 3).

In the second analysis, conformity was measured as difference: for each of the 22 critical questions, there was a difference between the number of answers matching the incorrect and fabricated majority answers showed to the conformity group and the number of participants giving the same answers in the control group. χ^2 was used to determine if the frequency of answers matching the manipulation in the conformity group differed from the frequency of the same answers in the control group.

According to the results of the χ^2 test, there was a significant difference between the experimental and control group for all the questions, except three of them (q15, q23 and q30). So the experimental group conformed much more than the control group.

Both analyses (mean scores and χ^2 test) support hypothesis 1: the conformity group conformed to what they were led to believe was the answer of the majority of the participants to a greater degree than the chance of participants in the control group giving the same incorrect answer. The results show that conformity behavior occurs in an internet-based context. Although it is not as great as what Asch observed, it is still remarkable.

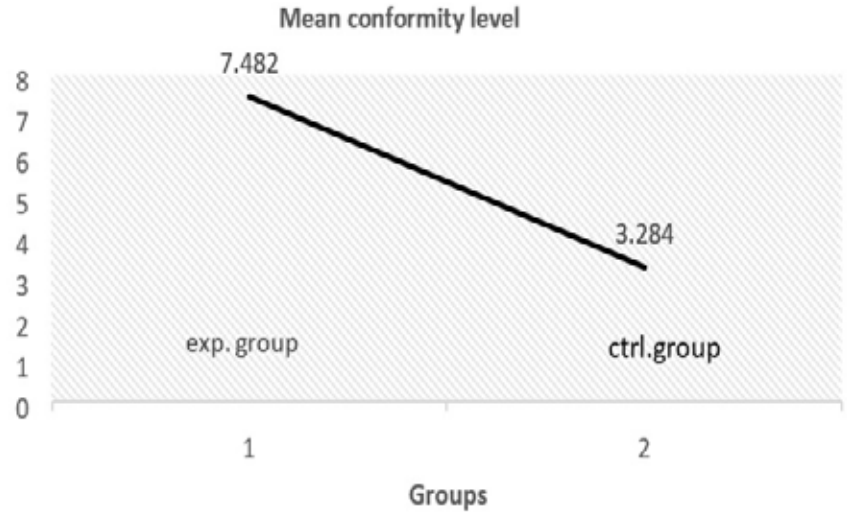


Figure 3: Comparing mean conformity levels between experimental and control groups.

3.1 People with high moral competence are less likely to succumb to conformity pressure

According to this hypothesis, we expected a significantly lower level of conformity behavior among the high C-score group.

The mean score for moral competence in the experimental group (M=20.552,

SD=16.379, N=112) was almost compatible with the control group (M=20.821, SD=15.71, B=96). The results reveal that participants with moral competence with C (for moral competence) higher than 20 (C >20) show less conformity than participants with lower than 20 (C <20). Moral competence and conformity behavior were negatively correlated (r = –0.183). Higher C-scores are compatible with a decrease in conformity behavior in participants (Figure 4). As can be seen in Figure 4, the correlation is rather small. This occurs maybe for the following reason: the pressure to conform was not high enough to show a wider difference in the conformity level for different C-scores, but as can be seen in Figure 2, these two variables are still negatively correlated.

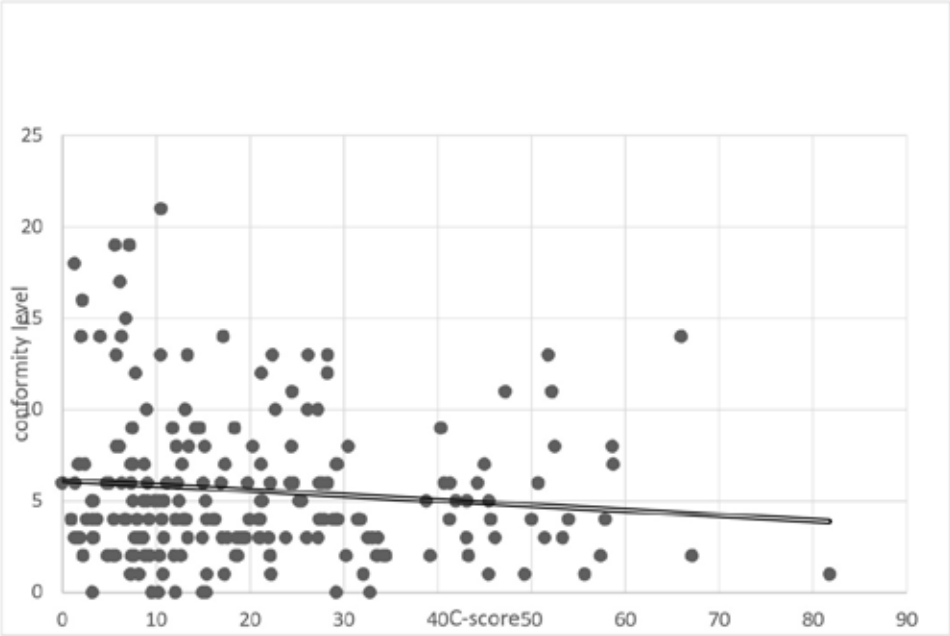


Figure 4: Participants' conformity level with different C-scores.

4. Discussion

The main purpose of this study was to investigate the impact of moral competence on conformity behavior in a non-face-to-face situation. As the results indicate, there is a weak but negative correlation between moral competence score and conformity behavior. When compared to Asch's results, there seems to be a great decrease in conformity behavior as presented by the participants. We assume the effects of deindividuation may be one of the main reasons for this decrease, as Rosander and Eriksson (2012) also reported. But the evident decline in mean conformity level is consistent with other previous studies of online conformity (Rosander & Eriksson 2012; Jim & Park 2011; Cinnirella & Greene 2007). Comparing to Rosander and Eriksson (2012), the mean conformity of this study is also lower (7.4% compared with 13.0%), which may relate to the differences in culture, experimental design and the sample size. As there are few studies in this area, further replications of this study are needed with a larger sample and with some minor changes in the experimental design, in a manner that will increase the amount of social pressure.

Some suggestions in this regard:

• The questions which were chosen for the experimental group could be selected more accurately in terms of difficulty, so that the pressure to conform increases; thus a standard pilot study may be needed prior to the experiment.

• Any measure that decreases the deindividuation effect would boost the conformity effect; for example, if participants were told that their answers would be displayed publicly on the website after the test, with their own e-mail address visible, it may influence their behavior to some extent.

In accordance with Hypothesis 2, the results show that moral competence can play a role in conformity behavior. In other words, conformity behavior may not be as fixed as Asch suggested. Rather, moral competence seems to be a mediator variable which affects the level of conformity. As the present research was a preliminary study in this area, a more accurate experimental design with a larger sample size is needed to address this subject in a more reliable manner.

On the other hand, in his conformity experiments Solomon Asch showed that more often than not people value others' opinions more than their own, even when the answers provided by the majority are obviously wrong. Explanations for this behavior, such as participants turning to the group for guidance and avoiding social isolation, have been discussed in earlier research, but all of these factors can be considered environmental, that is, a changing social environment seems to be the only factor that may affect the level of conformity behavior.

However, in this study an "internal variable" has been found which seem to be stronger than other previous ones. Moral competence is an ability which facilitates individuals to act on the basis of their own opinions, regardless of how intense the pressure is on the outside. As we could see in the results, people with higher C-scores were less likely to accept the others' answers, even when the pressure to conform was high due to the fabricated majority answers. In contrast, people with low or medium C-scores conformed to almost every critical question. In addition, there were few individuals in Asch's experiments who never conformed during the test trials, which may be another hint for researchers to focus more on internal reasons rather than environmental ones. Perhaps there is a need for a new model of behavior, an internal model which includes such characteristics as moral competence interacting with conformity behavior. One of the main goals of this study was to find a new way of understanding the influence of social media on people's behavior, since they have recently become a major issue in Iranian society. As Lind (2016) states, moral competence is an educational concept which can be learned through development. According to his definition, conformity behavior seems to change through learning moral competence. This could be one of the new main policies of educational systems all around the world, especially in developing countries like Iran.

5. Conclusions

The present study aimed to see if and how people conform in an online setting and how moral competence would affect people's conformity behavior in this context. The results show that people do conform in this situation; that is, even people who are online by themselves are still under the pressure of the majority to produce conformity behavior. This is consistent with earlier research (Rosander & Eriksson 2012; Jim & Park 2011; Cinnirella & Greene 2007). When compared to Asch`s line judgment task, the mean conformity in this experiment is lower (36.8% compared with 7.4%), which is because of many differences in experimental design, especially the online form of experiment which is significantly different from face-to-face situation.

Overall, although this result was rather weak in comparison with other face-to-face experiments, it still could indicate that people's level of moral competence could affect their social behavior, even in an online platform. In other words, people with higher moral competence, which can be acquired through education, are less prone to showing conformity in conditions characterized by social pressure.

References

Asch S. E. 1955. "Opinions and Social Pressure," *Scientific American* 193(5):31–35.

Asch S. E. 1956. "Studies of Independence and Conformity. A Minority of One Against a Unanimous Majority," *Psychological Monographs* 70(9) (whole No. 416):1–70.

Asch S. E. 1958. "Effects of Group Pressure upon the Modification and Distortion of Judgements," in E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in Social Psychology* (pp. 174–183) *New York: Holt, Rinehart, and Winston.*

Baumeister R. F. 1982. "A Self-presentational View of Social Phenomena," *Psychological Bulletin 91*(1):3–26.

Bond R. & Smith P. B. 1996. "Culture and Conformity: A Meta-analysis of Studies Using Asch's Line Judgment Task," *Psychological Bulletin* 119:111–137.

Chen Y., Harper F. M., Konstan J., & Li S. X. 2010. „Social Comparisons and Contributions to Online Communities: A Field Experiment on Movie Lens," *American Economic Review* 100:1358–1398.

Cinnirella M. & Green B. 2007. "Does 'Cyber-conformity' Vary Cross-culturally? Exploring the Effect of Culture and Communication Medium on Social Conformity," *Computers in Human Behavior* 23:2011–2025.

Deutsch M. & Gerard H. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgement," *Journal of Abnormal and Social Psychology* 51:629–636.

Goeree J. K. & Yariv L. 2015. "Conformity in the Lab," *Journal of the Economic Science Association* 1(1):15–28.

Janes L. M. & Olson J. M. 2000. "Jeer Pressure: The Behavioral Effects of Observing Ridicule of Others," *Personality and Social Psychology Bulletin* 26*(*4):474–485.

Kim J. & Park H. S. 2011. „The Effect of Uniform Virtual Appearance on Conformity Intention: Social Identity Model of Deindividuation Effects and Optimal Distinctiveness Theory," *Computers in Human Behavior* 27:1223–1230.

Kohlberg L. & Candee D. 1984. "The Relationship of Moral Judgment to Moral Action," in W. M. Kurtines & J. L. Gewirtz (Eds.), *Morality, Moral Behavior, and Moral Development* (pp. 52–73). New York, NY: Wiley.

Lind G. 1978. „Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts" [How does one measure moral judgment? Problems and alternative ways of measuring a complex construct], in G. Portele (Ed.), *Sozialisation und Moral* (pp. 171–201). Weinheim: Beltz.

Lind G. 1989. "Measuring Moral Judgment: A Review of 'The Measurement of Moral Judgment' by Anne Colby and Lawrence Kohlberg," *Human Development* 32(6):388–397.

Lind G. 2000 (Orig. 1984). *Content and Structure of Moral Judgment* (2nd Corrected Edition). Doctoral Dissertation, University of Konstanz.

Lind G. 2016. *How to Teach Morality: Promoting Deliberation and Discussion Reducing Violence and Deceit*. Berlin: Logos Verlag GmbH.

Milgram S. 1963. "Behavioral Study of Obedience," *The Journal of Abnormal and Social Psychology* 67(4):371–378.

Rosander M. & Eriksson O. 2012. "Conformity on the Internet – The Role of Task Difficulty and Gender Differences," *Computers in Human Behavior* 28(5):1587–1595.

Saeidi-Parvaneh S. 2011. *Moral, Bildung und Religion im Iran: Zur Bedeutung universitärer Bildung für die Entwicklung moralischer Urteils-und Diskursfähigkeit in einem religiös geprägten Land* (Doctoral Dissertation, University of Konstanz).

Stang D. J. 1976. "Ineffective Deception in Conformity Research: Some Causes and Consequences," *European Journal of Social Psychology* 6:353–367.

Zimbardo P. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*. New York: Random House.

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #6, pp. 84-94. DOI:10.14746/eip.2021.1.7

Panic and the Lack of Moral Competence. How We Can Help to Prevent Panic Pandemics

Georg Lind

(University of Konstanz; Germany; georg.lind@uni-konstanz.de)
ORCID: 0000-0002-2235-5465

Abstract: Often we have to decide on difficult problems and conflicts. For this, a certain level of moral competence is needed, in order to solve them as quickly and adequately as possible. Otherwise these problems and conflicts can overwhelm us, triggering a feeling of fear and panic, and making us react too slowly or inadequately, or both. Fear and panic can make us ignore problems and conflicts, attempt to "solve" them through brute force or deceit, or declare them to be beyond our responsibility and let an authority decide what to do. Often such makeshift solutions seem to work, but, more often, they have damaging effects. Therefore, society tries to curb criminal and anti-democratic activities through coercion, that is, through laws, law-enforcing institutions, and correction facilities – at high costs, and often with little efficacy.

In this article I show that such coercion would not be needed if we gave all citizens an opportunity to develop their ability to solve conflicts and problems through thinking and discus-sion. Moral competence would immunize us against fear and panic, and thus also against immoral practices. Moral competence is not inborn in us, and it does not develop unless it is fostered through proper learning opportunities. Therefore, if we want to live together peace¬fully in a democratic society, we need to provide proper learning opportunities for everyone, not only of a few people. If the masses are infected by panic, a few rational people cannot stop this pandemic.

Keywords: Moral competence; panic; pandemics; panic pandemics.

Fear and Panic

Why and how are fear and moral competence connected? Let us start by looking at the nature of fear and its extension – panic. Fear is a mixed blessing. On the one hand, fearful reactions can save us when we encounter a threat, especially when a quick reaction is required. But it can also harm or even kill us when it prevents us from fully understanding the threat and leads us to make the wrong decisions. If fear is so strong that it prevents us even from reflecting on, and learning from, our decisions, then we are dealing with panic.

What happens in our brain when we feel fear? When we encounter a threat, a snake for example, our body is immediately activated in order either to fight the threat or to flee from it. Both reactions may save us from being killed by the snake. But both can also make things worse for us. If we decide to fight, the snake might bite us. If we decide to flee, we may fall into an abyss and break our neck.

LeDoux (1994) has studied what happens in our brains when we see, hear, or smell a threat, and when this triggers fear in us. When the information from our senses is recognized by the brain areas for seeing, hearing and smelling, they send a signal to our limbic system in the lower part of the brain. This area comprises the amygdale, the seat of our emotions, the thalamus, which relays information to the other parts of the brain, the hypothalamus and the hippocampus, the locations of homeostasis and of memory consolidation, respectively. The hippocampus seems to store our encounters with threats. That is, our brain recognizes situations as threats only when we have stored respective threatening experiences (e.g., with snakes) or have stored warnings from trusted authorities (parents, teachers, media, government, etc.) or peers. As parents, often we have to warn our children to stay away from harmful things like a hot oven because they would touch it fearlessly and burn themselves. Fear is more often learned from other people than from experience, though the latter may be more lasting.

Fear is triggered not only by immediate dangers like wild animals, thieves or explosions but also when we encounter challenging tasks at school, at the workplace, in the political domain, or when our health is at stake. In those instances, we have to answer difficult moral questions. To which alternative should we give priority? Should we try to solve the task in a math test or should we just guess, or should we copy the answer from others? Should we tell a patient that we do not know how to help him or should be prescribe some pills to please him? Should we elect the candidate who would improve my life conditions or the candidate who is best for the country? To give a more current example: Should we get vaccinated against a certain virus, or should we distrust the safety and effectiveness of the vaccination? Like other questions, this decision also triggers difficult moral questions that may overburden our ability to solve dilemmas (see Figure 1 below) and thus create fear in us, or even panic (this example is taken from Engelbrecht & Köhnlein 2020).

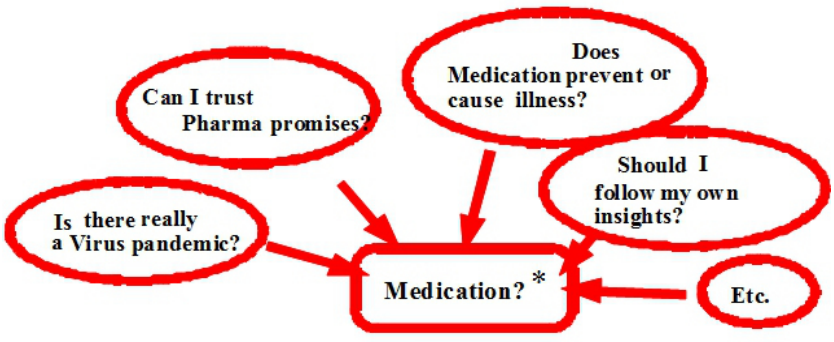


Figure 1: What can make a decision so difficult (examples for such decisions are found in the article by Engelbrecht and Köhnleim (2020).

Panic Is Shared Fear

If several people, a whole community or a whole nation feel the same kind of fear – like fear of a common foe or fear of a virus pandemic – we can refer to this as panic. In general, we feel panic less intensively than fear, because when everybody shares the same fear, it feels more normal and acceptable. At the same time, it may have a stronger impact on our behavior and on society, because we assure each other that the cause for the panic still persists, even long after it has gone. Panic cripples our thinking and discussion more than individual fear, because there is hardly anyone left who may stimulate our thinking and start a discussion. Panic tends to become chronic because we reinforce each other's fears, and also because, after a while, the brain shrinks parts of the frontal lobe when they are not used for a longer time, just as the muscles of our legs shrink when we do not use them. Even if we want to think again, we cannot think as well as before, because our brain lacks the necessary "hardware," that is, dendrites and synapses. Because we tend to see our panic as normal, it hardly ever shows up in surveys. However, it can be observed in our behavior. There are some sure signs of a pathological state of panic: When we suffer from panic, we tend to avoid thinking and talking about our object of fear at all. If someone happens to mention it, we tend to end any conversation or change the topic. Blinded by our panic, we think that these people are sick, crazy, or prejudiced against the truth, and had better shut up or be locked up so that they cannot bother us anymore. If we are arguing, our arguments lack any logic. Panic makes us contradict ourselves. A newspaper wrote about the coronavirus: "The infection rate is still rising. But the death rate remains very low. This is because people do not get sick." Sure, if nobody gets sick, the death rate is unlikely to increase. But why do people not get sick if the virus is so deadly? Is it not so dangerous after all? But if the answer is yes, we would not need to panic anymore! Does this thought let us bias our logic? Panic, it seems, wants to keep itself alive by allowing us an illogic which under normal conditions we would never accept.

When in a panic, we use anecdotes and examples for defending our truth, rather than scientifically based numbers, tables and graphs. When a close friend has become sick with a certain virus, we are sure that this is proof enough of the existence of a pandemic. Even if there are many statistics which show that the virus is not a "killer" but a normal flu, our panic leads us to believe more in single cases than in the statistics. When in a panic, we also tend to discard other, potentially greater dangers for our health, like polluted air and water, or false medication. This is because our brainstem can only handle singles causes and does not allow our frontal lobe to bother us with alternative causes as explanations for a threat.

The same is true for the means which we consider for fighting the threat. Again, when panic blocks our ability to think, our options are confined to only one means. To use our example, we are happy that our government tells us that there is only *one* way to protect us from the virus, namely through vaccination, in spite of the fact that many experts

believe that the search for a vaccine must fail, and that our immune system (the "T-cells") can protect us more effectively[1]. When we are in a panic, we cannot weigh the arguments for and against vaccination because this would overburden our thinking ability. Hence, we choose the sources of information which we trust more. In normal times this would be science. But in a state of panic, we tend to trust the more powerful source.

This shift of focus from the message to the messenger is typical for panic behavior. In a state of panic we can ignore discomforting information. But we cannot ignore the people who confront us with it. We think of them as enemies who do not deserve respect and civil conversation, but defamation ("ill-minded liars," "misled dumb-heads," "conspiracy theorists")[2].

As Gustave Le Bon (1897) showed 120 years ago in his seminal book *The Psychology of Masses*, panic is often used for political repression. Politicians are tempted to use our panic for their purposes because panic paralyzes not only our thinking but also our willingness to ask questions and fight for our rights. We let them restrict our civil rights, e.g., our rights of free speech and of free movement, because when we are in a state of panic, we find it hard to make use of our freedom. We should remember that these threats do not have to exist in order to cause panic, but need only be communicated by sources that we trust, like our favorite newspaper, TV program, or politician.

Besides politicians, other actors, like the media and businesses, might also be tempted to exploit our panic. When we perceive a danger, we buy more newspapers and watch more TV. In times of international tensions, we expect our government to order more weapons and security devices. When we are afraid of a virus pandemic, we demand distancing rules, face masks, mass testing and mass vaccination. In that case, the beneficiaries are many: manufacturers and distributors of masks, medicines, tests and vaccines, and their shareholders. What reasons should these actors have to declare the end of the panic when the statistics show that there was only a common flu and that this has already ended? These actors may even be tempted to keep our panic alive with (fake) information about an allegedly ongoing pandemic, like we did to the girls when we were young: "Watch out, the spider is still on your back."

Eventually panic feeders may find themselves caught in a vicious circle, namely when they cannot stop the panic-pandemic anymore, even when its costs have become unbearably high. Now citizens might believe so deeply in the existence of a pandemic that they demand the lockdown be continued. This explains why panic can persist even when the damage done by the panic far exceeds the damage caused by the original danger. Sometimes, a panic will end in a fatal disaster (LeBon 1897).

1 For extensive information and scientific articles on this issue see the CHD's website: https://childrenshealthdefense.org/.

2 For example, Dr. Wolfgang Wodarg, who has decades of experience and competence as a physician, epidemiologist and health politician, believes that the current corona "pandemic" is just an ordinary flu. For this belief, which is supported by many scientific studies, he is vilified in television programs and was thrown out of an association of which he was a member of the board. His videos on the internet were blocked several times.

Fostering People's Moral Competence Can Prevent Panic

Yet there is hope, at least for the future. Studies show that not all of us become panic-stricken and behave irrationally when we are faced with a (real or a communicated) threat. Many of us are able to weigh alternatives and to discuss controversial matters with others, and to make our decisions on the basis of our moral principles. For these people, difficult problems and conflicts are no threat and do not cause them to panic. When we understand why they can handle problems and conflicts under pressure and do not show signs of panic, we may be able to say how we can help the others to achieve the same.

The key to this understanding is *moral competence*, the technical term for our ability to solve problems and conflicts, which involve moral principles, through thinking and discussion (Lind 2019a). Favorable learning opportunities are necessary to develop this competence (Schillinger 2016). Because people differ greatly with regard to the number and quality of such opportunities, they also differ widely with regard to their moral competence. If we had too few such opportunities, it is very likely that we will fail to solve problems and conflicts through thinking and discourse, and that we will have to use means which we consider immoral ourselves: ignorance, violence, deceit and blind obedience to an authority. The importance of moral competence may become clear when we look closely at some classical psychological experiments.

Stanley Milgram (1974) showed in his famous experiment that obedience to an authority can paralyze our moral conscience. The experimenter, who assumes the authority of a psychological researcher, instructs participants to give learners electric shocks when they made an error in a learning task. The shocks and the reactions of the victims are fake, but the participants are not aware of this. The findings show that most obey the instructions, although they cannot see the great pain that they cause. Milgram concluded from this finding that human behavior is fully under the control of external authorities, and that inner instances like a moral conscience cannot influence people's behavior.

Erich Fromm contradicted this interpretation. For him, Milgram's study actually shows the "presence of conscience in most subjects, and their pain when obedience made them act against their conscience" (Fromm 1973, 75). His interpretation is supported by the reports which the participants gave after the experiment was finished, and by fact that some stopped torturing the learners prematurely.

But why did they stop, and why not the others? What enabled them to do so? Milgram in advertently gives us a hint. He reports that participants with a higher level of education were more disobedient that those with a lower level (Milgram 1974, 234). Do resisters possess an ability which has been stimulated through their education? The answer seems to be yes, as Kohlberg (1984) has shown in an experiment similar to Milgram's. He not only recorded how obedient the participants were, but also assessed their moral competence.

Kohlberg created a clinical interview method for measuring moral competence. This scale is known as the Moral Stages. On the basis of my new understanding of moral competence and following a critical analysis of the clinical interview method (Lind 1989), I have developed an objective method of making moral competence visible in the pattern of people's responses to an experimentally designed test, the Moral Competence Test (see Lind 1978; 1981; 2019a). In terms of showing the power of moral competence in people's behavior, we can consider both methods as equally valid.

In his experiment, Kohlberg found that obedience to authority was indeed strongly correlated with moral competence. Of the participants with a high moral competence (Stage 5 "principled morality"), 75% resisted authority, compared to 13% in the group of lower moral competence (Kohlberg 1984). This finding suggests that if our moral competence is sufficiently developed, we can solve difficult problems and conflicts without the need to panic.

Franz-Josef Mansbart (2001) has shown that in fact participants with low moral competence scores need considerably more time for solving dilemmas than high scorers (r = -.36; my calculation, see Lind 2002).

Kristin Prehn and her colleagues were even able to find the main location of these processes, namely in the right dorso-lateral prefrontal cortex (DLPFC). We may call it the seat of moral competence. Whereas all the parts of our nervous system are more or less involved when we are confronted with a moral dilemma, the DLPFC is the busiest part of our brain. The correlation between the amount of activity (measured by the BOLD level) in that area and the C-score of the *Moral Competence Test*, was unusually high; r = -0,47. As in Mansbart's experiment, the participants with low moral competence needed much longer for deciding behavioral dilemmas than the participants with a high moral competence. Li *et al*. (2016), who ran a similar series of experiments, corroborated Prehn's finding.

More support for our hypothesis comes from studies of conformity behavior. Solomon Asch (1955) showed that most of us are ready to let others think for us. He showed that we often trust our own thinking less than the thinking of other people if they are more numerous. Psychologists call this phenomenon *conformism*. In his experiment, Asch asked the participants which of three lines on a sheet of paper had the same length as the line on another piece of paper. When the other participants, who had been prepared by the experimenter, gave the same wrong answer, most of the participants changed their correct answer in order to agree with the majority. Obviously, they trusted the opinion of the majority more than their own thinking. But, like Milgram, Asch forgot to ask why some of them resisted the temptation of conformity? Aida Mofakhami (2019) repeated Asch's experiment online, but also measures her participants' moral competence. Indeed, participants with high moral competence were less prone to the conformity effect. Her effect was weak, maybe because the social conformity pressure was mitigated by the online medium and, therefore, not as strong as in Asch's experiment with real groups of participants.

Sharon McNamee (1977) found that participants with low levels of moral competence were less likely to help people in distress. However, these participants were not less willing to help but, as they reported afterwards, they felt paralyzed by the conflicting thoughts and feelings that the helping situation had elicited in them.

The importance of moral competence for solving problems is also shown in the study by Beke Lenz (2006) on youths' drug consumption. She found that adolescents who faced severe life problems (e.g., losing a friend, getting of bad grades in school, suffering from their parent's divorce) tend to use drugs to calm down their feelings – but only when they had a low moral competence. These youths were not able to cope with their problems through thinking or through consulting with their parents, friends or teachers. Therefore, they had to fight the emotions which these conflicts caused through alcohol, smoking or other drugs. In contrast, participants with higher moral competence were apparently able to cope with their problems and, therefore, did not need to use drugs in order to calm their emotions.

These and many other experimental studies suggest (a) that moral competence really is a very important determinant of our behavior, and (b) that there seems to be a certain level of competence which must be attained if we want to solve our problems and conflicts successfully through thinking and thus avoid panic (Lind 2019a; 2019b). In terms of measurement, this critical level is marked by a C-score of 20.0 in the *Moral Competence Test* (see Figure 2). I should mention that this C-score is not a clear cut-off score but a statistical signpost. It should not be applied to individuals but only to groups of people.

If we are unable to develop this level of moral competence, we mostly feel overburdened by living in a free, democratic society. We feel permanent stress and easily panic when something unusual (like a virus pandemic) happens. In order to reduce this stress-feeling we try to "solve" problems and conflict in everyday life by ignoring them, or, if we cannot ignore them, by using violence and deceit, or, if this does not work, by letting authorities think for us and solve our problems.

Figure 2: The critical level for moral competence to become relevant for behavior.

When citizens lack even a minimum amount of moral competence, society must spend a big share of its budget for law-making, law-enforcement, judicial decision-making, correction facilities, making reparations to victims and fixing the damage caused by criminal behavior. Just imagine that a small misdemeanor like traveling by bus without

a ticket can result in a prison sentence of several months. Such a small damage of a few dollars can cost society several thousand dollars, in terms of punishing the transgressors. I believe that this money could be better spent on fostering their moral competence. So if we want to maintain and to improve our democracy, we must, above all, foster the moral competence of all citizens. This can neither be achieved by classical education nor by conventional civic education, as J. Westheimer rightly argues:

> In study after study, we come to similar conclusions: the kinds of goals and practices commonly represented in curricula that hope to foster democratic citizenship usually have more to do with voluntarism, charity, and obedience than with democracy. In other words, good citizenship to many educators means listening to authority figures, dressing neatly, being nice to neighbors, and helping out at a soup kitchen - not grappling with the kinds of social policy decisions that every citizen in a democratic society needs to learn how to do (Westheimer 2015, 472).

If they want to serve our democracy well, our schools have to foster our children's ability to think for themselves, to trust their own thinking, and to discuss controversial issues with opponents. That is, in order to lift the overall level of moral competence in our societies to the needed minimum level, its schools have to provide their students with sufficient opportunities for using and training their moral competence, and to participate in exchanges with others (Figure 3).

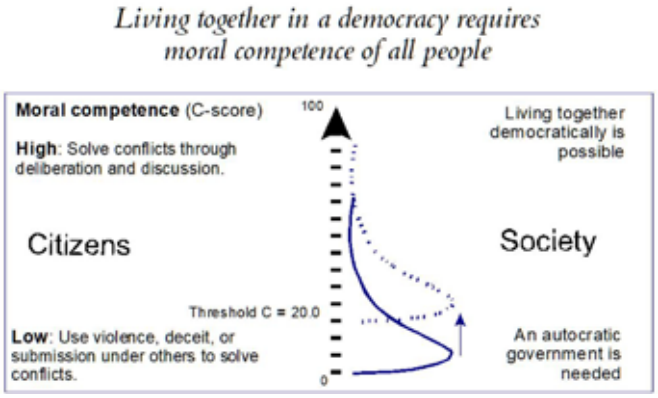


Figure 3: The current level of moral competence in most societies (solid line) is too low for living together peacefully in a democracy. We must it above the threshold (dotted line).

This implies a big shift in teaching methods. Students' thinking and discussions must not be guarded and graded by an authority (the teacher) but must follow moral-democratic principles. Actually, it must follow only one basic principle, namely the free speech-principle: everyone is allowed to say anything he or she wants, but must never qualify (praise or blame) any other participant. Having over twenty years of experience of working with this simple principle, I feel entitled to say: It works! Never has any participant transgressed. All I needed to do was to announce this principle and "threaten" to remind anyone who would break it. I never had to remind anyone.

This principle is the center piece of the *Konstanz Method of Dilemma Discussion*®

(KMDD), which I have developed on the basis of Blatt and Kohlberg's (1975) method of dilemma discussion (Lind 2019a; Reinicke 2017). The KMDD has been tested and evaluated many times (Lind 2019a). The KMDD is a very effective and yet not very complex method for the promotion of moral competence. However, the KMDD's effectiveness depends on the quality of the training of the teachers who use it. Hopefully, institutions of higher education will begin to offer such training. I would gladly assist.

Conclusions

In this article I have shown that we do not need to coerce people into democracy, which is rather paradoxical, if we foster everyone's ability to solve the conflicts and problems, which democracy inevitably confronts us with, through thinking and discussion. This moral competence would immunize people against fear of freedom and against panic, and thus also against immoral practices. Moral competence is not inborn and does not develop unless it is fostered through proper learning opportunities (Lind 2000, 2020; Schillinger 2006). Therefore, if we want to live together peacefully in a democratic society, all people must be provided with sufficient opportunities for using their moral competence. This is the most honorable task of schools in a democracy.

References

Asch S. E. 1955. "Opinions and Social Pressure," *Scientific American* 193(5):31–35.

Blatt M. & Kohlberg L. 1975. "The Effect of Classroom Moral Discussion Upon Children's Level of Moral Judgment," *Journal of Moral Education* 4:129–161.

Bundesverwaltungsgericht 2005. „Urteil des 2. Wehrdienstsenats vom 21. Juni 2005," *BVerwG* 2 WD 12.04. URL: https://www.bverwg.de/210605U2WD12.04.0 (last accessed on March 7, 2021).

Fromm E. 1973. *The Anatomy of Human Destructiveness*. New York: Fawcett Crest Books.

Engelbrecht T. & Köhnlein C. 2020. „Die Medikamenten-Tragödie" [The drug tragedy]. *Rubikon* September 29, 2020. URL: https://www.rubikon.news/artikel/die-medikamenten-tragodie (last accessed on March 8, 2021).

Gullone E. 2000. "The Development of Normal Fear: A Century of Research," *Clinical Psychology Review* 20(4):429–451.

Hemmerling K. 2014. *Morality Behind Bars – An Intervention Study on Fostering Moral Competence of Prisoners As a New Approach to Social Rehabilitation*. Frankfurt am Main: Peter Lang Editions.

Kohlberg L. 1964. "Development of Moral Character and Moral Ideology," in M. L. Hoffman & L. W. Hoffman (Eds.), *Review of Child Development Research.* Vol. I (pp. 381–431). New York: Russel Sage Foundation.

Kohlberg L. 1984. *The Psychology of Moral Development.* Vol. 2: *Essays on Moral Development*. San Francisco: Harper & Row.

Le Bon G. 1895. *The Psychology of Masses*. Cardiff, UK: Sparkling Books.

LeDoux J. E. 1994. "Emotion, Memory and the Brain," *Scientific American* 270:32–39.

Li X., Yang J., Li P., & Li H. 2016. "Individual Differences in Moral Judgment Competence Are Related to Activity of the Prefrontal Cortex When Attributing Blame to Evil Intention," *Social Neuroscience* 11(4):438–448.

Lind G. 1978. „Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts" [How does one measure moral judgment? Problems and possibilities of measuring a complex construct], in G. Portele (Ed.), *Socialisation und Moral* (pp. 171–201). Weinheim: Beltz Verlag.

Lind G. 1989. "Measuring Moral Judgment: A Review of 'The Measurement of Moral Judgment' by Anne Colby and Lawrence Kohlberg," *Human Development. Essays on Moral Development* 32:388–397.

Lind G. 2000. "The Importance of Responsibility-taking Opportunities for Self-sustaining Moral Development," *Journal of Research in Education* 10(1):9–15.

Lind G. 2002. *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung*. [Can morality be taught? Research findings from modern moral psychology]. Berlin: Logos Verlag.

Lind G. 2019a. *How to Teach Moral Competence. New: Discussion Theater*. Berlin: Logos.

Lind G. 2019b. "Why We Need to Foster Moral Competence and How." Presentation at the 13th International Symposium "Moral Competence: Its Nature, Relevance and Teachability," University of Konstanz, July 24-25, 2019.

Lind G. 2020. "Moral Competence: What It Means and How Accountant Education Could Foster It," in M. Pinheiro & A. O. Costa (Eds.), *Accounting Ethics Education* (pp. 155–174). London: Routledge.

Mansbart F.-J. 2001. *Motivationale Einflüsse der moralischen Urteilsfähigkeit auf die Bildung von Vorsätzen*. Unpublished master thesis. Dept of Psychology, University of Konstanz, Germany.

Milgram S. 1974. *Obedience to Authority. An Experimental View*. London: Tavistock Publications.

Mofakhami A. 2019. "The Effect of Moral Competence on Online Conformity Behavior." Presentation at the 13th International Symposium "Moral Competence: Its Nature, Relevance and Teachability," University of Konstanz, July 24-25, 2019.

Prehn K. 2013. "Moral Judgment Competence: A Reevaluation of the Dual-Aspect Theory Based on Recent Neuroscientific Research," in E. Nowak, D. E. Schrader, & B. Zizek (Eds.), *Educating Competencies for Democracy* (pp. 9–22). Frankfurt am Main – Bern – Brussels – New York: Peter Lang Editions.

Reinicke M. 2017. *Moral Competence Reloaded*. Private publication. Available from martina.reinicke@outlook.de.

Schillinger M. 2006. *Learning Environments and Moral Development: How University Education Fosters Moral Judgment Competence in Brazil and two German-speaking Countries.* Aachen: Shaker Verlag.

Westheimer J. 2015. "Teaching for Democratic Action," *Educação & Realidade* 40(2):465–483.

Further sources on moral-democratic competence. URL: https://www.uni-konstanz.de/ag-moral/home-e.htm

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #7, pp. 95-103. DOI:10.14746/eip.2021.1.8

The Impact of the Moral Foundations Arguments on Early Adolescents

Kamila Stastna

(independent scholar; Pirna; Germany; kamila.stastna@seznam.cz)
ORCID: 0000-0003-0867-7393

Abstract: The empirical research reported in this article is based on the Moral Foundations Theory proposed by J. Haidt. Objectives. The author examines the impact of moral foundations arguments on early adolescents' moral judgments regarding violating moral rules and explores gender-related differences between moral foundations preferences. Method. The effect of moral foundations arguments was measured by a newly developed meta-ethical position test (MEPT). The MEPT consists of a pretest questionnaire, treatment by moral foundations arguments, and a posttest questionnaire. The sample contained 178 early adolescents from the Czech Republic (84 females and 94 males). The influence of the moral foundations arguments was analyzed by comparing the pretest with the posttest. Results. 91% of teenagers changed their moral judgment due to confrontations with the moral foundations arguments. A Wilcoxon signed-rank test found that the moral foundations arguments were significantly relevant, since the P-value was lower than 0.001. The Mann-Whitney U test revealed the importance of the gender aspect: P-value care equals 0.01 and liberty 0.01. Girls have a preference for care foundation (21% more than boys), while boys tended to liberty (27 % more than girls). It seems that moral foundations arguments strongly change early adolescents' moral judgments and can be practically applied as a valuable platform for early adolescents' moral development.

Keywords: Moral Foundations Theory; moral judgment; moral competence; moral foundations argument; gender; moral development.

Introduction

Why do some arguments change our moral judgment and others not? What gives them the special power to convince us? Are some types of arguments universal? Certain studies claim that universal arguments do not exist, and according to them morality is relative (Prinz 2007; Sargissian 2014; Quintelier 2013); others suppose morality is objective and believe that universal arguments exist (Nichols 2004; Goodwin & Darley 2012). This study is based on the Moral Foundations Theory (Haidt 2012) and aims to examine if arguments connected to moral foundations change early adolescents' moral

decisions. In addition, the second aim is to investigate if gender makes a difference to moral foundations preference. Some studies do not reveal any significant differences between females and males (Krebs *et al.* 1994; Jaffee & Hyde 2000), whereas others do (Benhabib 1985; Friesdorf 2015). This research uses a meta-ethical position test (MEPT) to measure the willingness to violate basic moral rules and the differences between the moral foundations preferences of males and females. A deeper understanding of moral foundations arguments can help teachers motivate their students and develop their morality naturally.

Theoretical Background

The Moral Foundations Theory was developed by the moral psychologist Jonathan Haidt and his colleagues, who investigated cultural differences regarding moral judgments and consequently found that a few moral values influence people worldwide. In conclusion, they suppose these values are innate; they develop gradually as a reaction to the human being's environment and are characterized as instinctive emotional dispositions. These fundamental values are labeled as moral foundations and comprise care/harm, liberty/oppression, fairness/unfairness, loyalty/betrayal, authority/subversion, and purity/disgust. However, moral foundations are, according to Haidt, universal; people just have different sensitivities towards them. Liberal-oriented people are more sensitive to care, liberty, fairness, and conservatives consider all categories equally (Haidt 2012). The preferences of moral foundations were measured by a self-report instrument, namely the Moral Foundation Questionnaire (MFQ), which has been examined worldwide. For example, a study involving 1,645 participants confirmed the validity of the MFQ in France (Métayer & Pahlavan 2014). Likewise, a study conducted in New Zealand also affirmed its applicability. This research comprised 3.994 people and focused on the factor structure of the MFQ, which used confirmatory factors to perform an independent test of the MFQ (Davies, Sibley, & Liu 2014).

Furthermore, the Moral Foundations Theory has been examined by neuroscience. Lewis and his colleagues scanned the brains of 70 young, healthy adults. According to their findings, people who have a preference for care and fairness were associated positively with the left dorsomedial PFC volume and associated negatively with bilateral precuneus volume. People who tended towards authority, loyalty, and purity showed an association to the bilateral subcallosal gyrus and the left anterior insula volume (Lewis *et al*., 2012). The influence of moral foundations regarding children and adolescents was indirectly investigated by Nucci, Turiel, and Roded, who focused primarily on moral objectivism. Their experiment involved 167 children and adolescents between 8-16 (17) years old, and showed that participants considered unconflicted situations, i.e., hitting, stealing, and not helping, as objectively wrong, whereas conflicted situations were judged differently in terms of age. For example, the participants 10-14 (11) years old were more likely than

others to follow and agree with the nonprosocial choice, i.e., not helping someone in need (Nucci *et al.* 2017). According to some researchers, understanding care, welfare, and fairness develop very early in comparison with the sense of justice. The latter develops later and associates with reasoning about mathematical and physical problems (Damon 1975; Smetana et *al.* 2014).

The first pioneer who explored the gender aspect in the context of moral orientation preference is the well-known British moral psychologist Carol Gilligan (1982). She investigated the Kohlberg moral typology and found that women are attached to care; men to justice. Studies dealing with gender differences have predominated from that time. For example, Jaffee and Hyde (2000) conducted a meta-analysis from over one hundred studies; nevertheless, the results revealed only slight differences regarding care and justice. Accordingly, the study of Graham *et al.* (2011), which contained 34,476 adults and used MEQ, also found slight differences between the genders. Females scored higher in purity, care, and fairness, and males in authority and loyalty. In contrast, Niazi *et al*. (2020) showed there is a significant difference regarding care associated with females. This research was focused on a Pakistani sample comprising 300 male and female participants, and the MEQ was applied. Similarly, the results of the recent study conducted by Atari *et al.* (2020) involving a large sample of 336,691 adults from 67 countries revealed that females scored consistently higher on fairness, purity, and care.

Hypotheses and Methodology

The studies mentioned above confirm that moral foundations influence moral judgment. This study presumes that the arguments embedded into the moral foundations questionnaire can motivate or even manipulate the decision-making process.

Hence, the first hypothesis posits that the moral foundations arguments impact the willingness to change early adolescents' moral judgment.

The second hypothesis deals with gender differences regarding moral foundations preferences. The recent studies cited above presented controversial results. Some researchers highlight the significance of care regarding females, while others reveal different moral foundations preferences. For this purpose, this hypothesis cannot specify any moral foundations preference and thus broadly posits that gender influences early adolescents' moral foundations preferences.

The impact of moral foundations arguments and moral foundations preferences were measured by a meta-ethical position test. The MEPT is divided into three parts: the pretest questionnaire, treatment, and the posttest questionnaire. The pretest questionnaire comprises the six different short moral stories dealing with the violation of basic moral rules (i.e., John intentionally beats up his classmate). Participants rate the extent to which they consider this behavior correct on the Likert scale between 0 – 4 (0 = strongly disagree and 4 = strongly agree). The second part of the MEPT presents

two sections. In the first section, respondents are asked to write motives for violating moral rules (i.e., Why might John intentionally beat up his classmate?). The second section presents the treatment by the moral foundation arguments aiming to justify violations of the moral rules (i.e., John intentionally beat up his classmate because his classmate had been intentionally spitting at him all day).

In the last part of the MEPT, the participants complete the posttest, which is the same questionnaire as the pretest. The MEPT is theoretically based on Moral Foundation Theory, and the construction of the six moral stories was influenced by the Konstanz Method of Dilemma Discussion (KMDD). These stories trigger thinking and emotion, the story's main protagonist usually has a name, and violates some standard moral rules, while the story vividly presents the daily life of adolescents (Lind 2019).

Sample and Procedure

Data were collected from six regular schools (in three different parts of the Czech Republic: north, south, and central regions) with the principals' cooperation. The sample comprised 178 early adolescents, 84 females and 94 males, aged 12 to 13. Pupils took the MEPT during class time in 2019, and the procedure took approximately 10 minutes.

Assessment

The relevance of moral foundations arguments was examined by a statistical comparison of pretest and posttest, applying a Wilcoxon signed-rank test, which compares two dependent samples of ordinal data (i.e., Likert scale). The importance of the gender aspect was investigated by a Mann-Whitney U test, which compares two independent samples (i.e., females and males).

Results

1) According to the results, 91% of early adolescents changed their moral judgments due to moral foundations arguments. The Wilcoxon signed-rank test revealed statistical differences between before and after treatment with all moral foundations arguments (see Table 1 and Figure 1)[1].

Moral foundation argument	Z-value	W-value	P-value	Result
Betrayal	8.4	1008	<0.00001	Significant impact
Subversion	7.3593	1397	<0.00001	Significant impact
Harm	8.4089	865.5	<0.00001	Significant impact
Unfairness	6.6287	1278	<0.00001	Significant impact
Oppression	7.7361	458.5	<0.00001	Significant impact

Table 1: Results of the Wilcoxon Signed-Rank Test.

1 The participants were also willing to violate moral rules in the pretest under the condition that they wrote an argument connected to moral foundations (i.e., John's case: self-defense, humiliation, self-defense, or revenge).

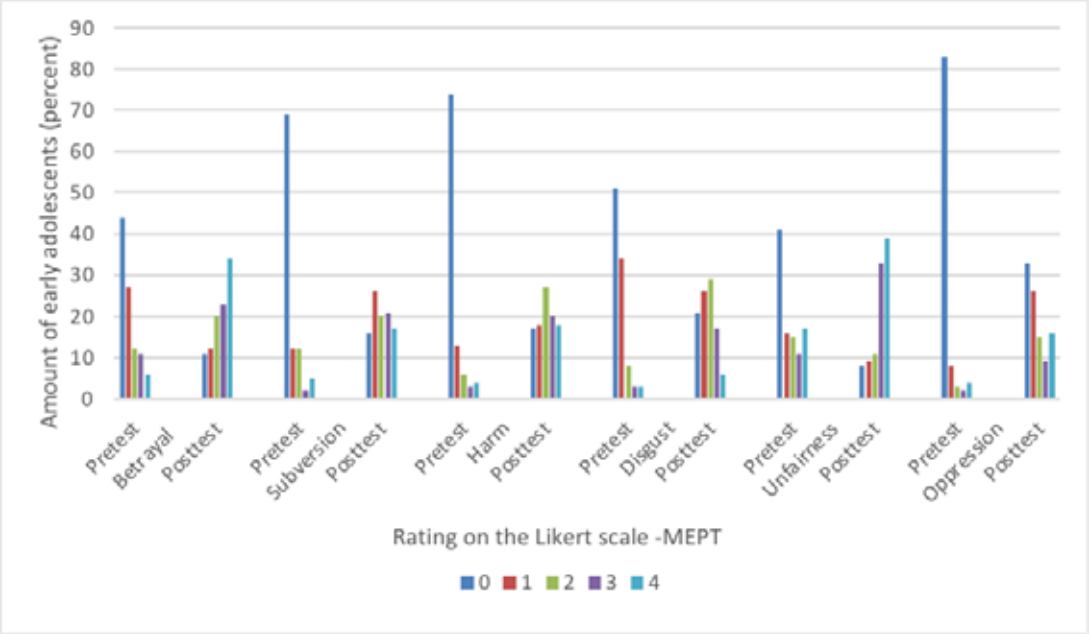


Figure 1: The impact of moral foundations arguments on moral judgment.

The Mann-Whitney U test revealed differences between females and males. A significant contrast was evident in oppression and harm, whereas others showed slight differences (see Table 2 and Figure 2).

Moral foundation	Z-score	U-value	P-value	Results
Unfairness	1.71938	3283	0.08544	not significant
Oppression	2.544	3039.5	0.01108	significant
Betrayal	1.45854	3409	0.1443	not significant
Harm	2.51609	3049	0.01174	significant
Subversion	0.79425	3591	0.42952	not significant
Disgust	0.80867	3464.5	0.41794	not significant

Table 2: Results of the Mann-Whitney U Test.

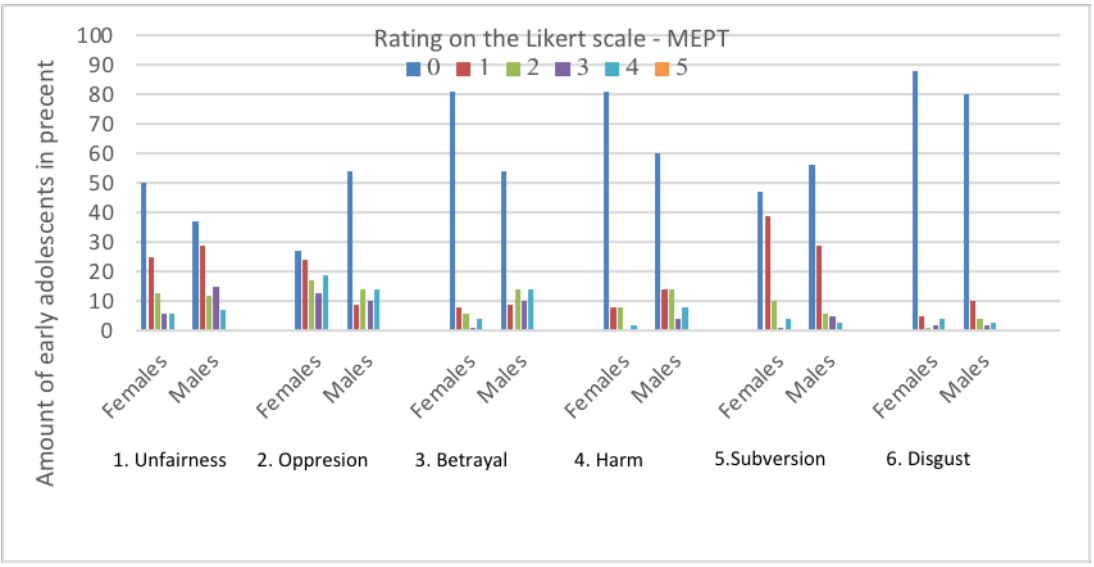


Figure 2: The gender impact on preferred moral foundations.

Discussion

The results reveal the moral foundations arguments strongly influenced the willingness to change early adolescents' moral judgments, P-value < 0,001 and support the finding of Haidt *et al.* (2012) and other studies (Yalçındağ *et al.* 2017; Davies, Sibley, & Liu 2014; Métayer & Pahlavan 2014; Zhang & Li 2015; Lewis *et al.* 2012) dealing with the universality of moral foundations. On the other hand, the findings of this research do not follow the studies reporting that people tend to moral objectivism. For example, Goodwin and Darley (2012) found that people consider some moral acts as universal truths. Similarly, the research conducted by Nucci *et al.* (2017) showed that participants across ages considered negative unconflicted situations, i.e., hitting, stealing, and not helping, objectively wrong. However, moral foundations arguments were regarded universally in the sense of motivation (activating the emotions and cognitive processes of moral judgments); understanding what is right and wrong was not universal, but rather relative to other aspects.

It appears that early adolescents have in their mind some scenario, some motives, some a priori reasons why an agent can act, and possibly this is based on individual experiences (i.e., boys found more motives than girls why John could intentionally beat up his classmate). These findings differ from Goodwin and Darley (2012) and Nucci *et al.* (2017), probably due to the controversial terminology of moral objectivism and applied methodology. In contrast to Goodwin and Darley (2012), this study was not conducted with adults. Modern neuroscience emphasizes some brain differences regarding adults' and adolescents' brains (Blakemore 2013).

The second findings refer to the gender aspect in connection to moral foundations preferences. The Mann-Whitney U test found statistical importance related to liberty and care. Fifty-four percent of boys, compared to 27% of girls, were convinced that oppressing was wrong. Eighty-one percent of girls were convinced that hitting classmates was wrong, contrary to 60% of boys. These results follow Carol Gilligan's theory and similar studies, showing that females are more likely to have a preference for care than males (Gilligan 1983; Friesdorf 2015). Gender differences in the context of moral foundations preferences are controversial in the research, and the reason for this should be further investigated (Graham *et al*. 2011; Jaffee & Hyde 2000; Atari *et al.* 2020; Niazi *et al.* 2020). This study has some limitations (i.e., conducted only in the Czech Republic, using nonparametric methods), and for this reason the findings cannot be generalized, and further studies are required.

Subsequent research could be conducted in different countries and religions; other studies could employ qualitative methods and apply neuroscience. Some studies could focus on practical innovation of moral education and investigate the impact of moral foundations arguments on mental perspective experiences. It could lead to the idea expressed by Gary Klein (2013). "I think helping people to arrive at insights isn't a

question of pushing the insights on the people or trying to explain it in words, as much as it is helping people to gain the experience so they can see the inconsistency for themselves, and then all of a sudden the mental model will shift naturally and easily," Klein (2013, 214) suggests.

Conclusion

This study dealt with moral foundations arguments and moral judgment regarding early adolescents. The findings reveal that the moral foundations arguments powerfully influenced the moral decision process and that gender impacted moral foundations preferences. Girls were attached significantly to care; boys to liberty. However, the moral foundations arguments appear to be universal regarding motivation (activating emotional and rational processes of moral judgments), the sense of understanding what is right and wrong seems to be relative to many other aspects (i.e., personal experiences). Consequently, applying the moral foundations arguments in a class can motivate students to see, think, and discuss moral issues more profoundly. Additionally, adolescents can learn to be more open and not condemn what they do not truly understand.

References

Atari M., Lai M., & Dehghani M. 2020. "Sex Differences in Moral Judgements Across 67 Countries," *Proceedings of the Royal Society B: Biological Sciences* 287(1937). DOI: 10.1098/rspb.2020.1201

Blakemore S. J. 2013. "The Adolescent Brain," in J. Brockman (Ed.), *Thinking: The New Science of Decision-making, Problem-solving, and Prediction* (pp. 115–131). New York: HarperCollins.

Benhabib S. 1985. "The Generalized and the Concrete Other: The Kohlberg-Gilligan Controverse and Feminist Theory," *PRAXIS International* 5(4):402–424.

Damon W. 1975. "Early Conceptions of Positive Justice as Related to the Development of Logical Operations," *Child Development* 46(2):301–312. DOI: 10.2307/1128122

Davies C. L., Sibley C. G., & Liu J. H. 2014. "Confirmatory Factor Analysis of the Moral Foundations Questionnaire," *Social Psychology* 45(6):431–436. DOI: 10.1027/1864-9335/a000201.

Friesdorf R., Conway P., & Gawronski B. 2015. "Gender Differences in Responses to Moral Dilemmas," *Personality and Social Psychology Bulletin* 41(5):696–713. DOI: 10.1177/0146167215575731.

Gilligan C. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, Mass. – London: Harvard University Press.

Goodwin G. P. & Darley J. M. 2012. "Why Are Some Moral Beliefs Perceived To Be More Objective Than Others?," *Journal of Experimental Social Psychology* 48(1):250–256. DOI: 10.1016/j.jesp.2011.08.006

Graham J., Haidt J., & Nosek B. A. 2009. "Liberals and Conservatives Rely on Different Sets of Moral Foundations," *Journal of Personality and Social Psychology* 96(5):1029–1046. DOI: 10.1037/a0015141

Haidt J. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon Books.

Jaffee S. & Hyde J. S. 2000. "Gender Differences in Moral Orientation: A Meta-analysis," *Psychological Bulletin* 126(5):703–726. DOI: 10.1037/0033-2909.126.5.703.

Klein G. 2013. *Seeing What Others Don't.* New York: Public Book Affairs.

Krebs D. L., Vermeulen S. C., Denton K. L., & Carpendale J. I. 1994. „Gender and Perspective Differences in Moral Judgement and Moral Orientation," *Journal of Moral Education* 23(1):17–26. DOI: 10.1080/0305724940230102

Lewis G. J., Kanai R., Bates T. C., & Rees G. 2012. "Moral Values Are Associated with Individual Differences in Regional Brain Volume," *Journal of Cognitive Neuroscience* 24(8):1657–1663. DOI: 10.1162/jocn_a_00239

Lind G. 2019. *How To Teach Moral Competence. New: Discussion Theater.* Berlin: Logos Verlag.

Métayer S. & Pahlavan F. 2014. "Validation de l'adaptation Française du Questionnaire des Principes Moraux Fondateurs" [Validation of the Moral Foundations Questionnaire in French], *Revue Internationale de Psychologie Sociale* 27(2):79–107.

Niazi F., Inam A., & Akhtar Z. 2020. „Accuracy of Consensual Stereotypes in Moral Foundations: A Gender Analysis," *PLOS ONE* 15(3). DOI: 10.1371/journal.pone.0229926.

Nichols S. 2004. "After Objectivity: An Empirical Study of Moral Judgment," *Philosophical Psychology* 17(1):3–26. DOI: 10.1080/0951508042000202354.

Nucci L., Turiel E., & Roded A. D. 2017. "Continuities and Discontinuities in the Development of Moral Judgments," *Human Development* 60(6):279–341. DOI: 10.1159/000484067.

Prinz J. 2007. *The Emotional Construction of Morals.* Oxford: Oxford University Press.

Quintelier K., Fessler D., Smet D. 2013. "The Moral Universalism-Relativism Debate," *Klesis Revue Philosophique* 27:211–262.

Sarkissian H., Park J., Tien D., Wright J. C., & Knobe J. 2011. "Folk Moral Relativism," *Mind & Language* 26(4):482–505. DOI: 10.1111/j.1468-0017.2011.01428.x.

Smetana J. G., Jambon M., & Ball C. 2014. "The Social Domain Approach to Children's Moral and Social Judgments," in M. Killen & J. G. Smetana (Eds.), *Handbook of Moral Development* (pp. 23–45). Hove, East Sussex: Psychology Press.

Yalçındağ B., Özkan T., Cesur S., Yilmaz O., Tepe B., Piyale Z. E., *et al*. 2017. "An Investigation of Moral Foundations Theory in Turkey Using Different Measures," *Current Psychology* 38(2):440–457. DOI: 10.1007/s12144-017-9618-4.

Zhang Y. & Li S. 2015. "Two Measures for Cross-Cultural Research on Morality: Comparison and Revision," *Psychological Reports* 117(1):144–166. DOI: 10.2466/08.07.pr0.117c15z5.

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #9, pp. 104-133. DOI:10.14746/eip.2021.1.9

Giving Moral Competence High Priority in Medical Education. New MCT-based Research Findings from the Polish Context

Ewa Nowak

(Adam Mickiewicz University in Poznań; Poland; ewanowak@amu.edu.pl)
ORCID: 0000-0002-5722-7711

Anna-Maria Barciszewska

(Karol Marcinkowski University of Medical Sciences in Poznań; Poland; ambarciszewska@ump.edu.pl)
ORCID: 0000-0002-7179-1474

Georg Lind

(Institute for Moral-Democratic Competence (IMDC) in Konstanz; Germany; georg.lind@uni-konstanz.de)

Kay Hemmerling

(Arbeit & Leben Institute in Leipzig; Germany; kayhemmerling@aol.com)

Sunčana Kukolja Taradi

(Medical Faculty, University of Zagreb; Croatia; skukolja@gmail.com)

Abstract: Nowadays, healthcare and medical education is qualified by test scores and competitiveness. This article considers its quality in terms of improving the moral competence of future healthcare providers. Objectives. Examining the relevance of moral competence in medico-clinical decision-making despite the paradigm shift and discussing the up-to-date findings on healthcare students (Polish sample). Design and method. N=115 participants were surveyed with a standard Moral Competence Test to examine how their moral competence development was affected by the learning environment and further important factors. Results. The sample allowed the identification of a regress in moral competence during students' pre-clinical curriculum, and progress during their clinical curriculum. A gender-related bias, a segmentation effect, and a pronunciation effect were noticed. Explanations. Scholarly literature usually reports a linear decrease of medical students' C-scores resulting from, e.g., competitive trends in education. We identified such trends in terms of gender-specific competitive tactics. Religious and ethical affiliations were discussed to explain the unexpected gender bias and the related segmentation and pronunciation effects. The findings can be regarded as predictive for similar developments in educational institutions regardless of cultural contexts as the sample examined in this article represents medical education in a country facing a transition from a non-competitive to competitive tertiary education model, and between presecular and monocultural to secular and pluralist social ethics.

Keywords: moral competence in healthcare professions; clinical decision-making; gender bias and moral competence; segmentation effect; pronunciation effect; Moral Competence Test; Polish medical education; competitive tertiary education.

1. Objectives

This article will firstly revisit the definition of moral competence and assess its relevance for healthcare professionals, bearing in mind the paradigm shifts observed in decision-making in medical and clinical contexts over the last decades[1]. Secondly, it will report on a recent MCT study conducted with Polish healthcare students to score their moral competence: this study uncovered the meandering developmental trajectory of their moral competence, a gender bias, and gender related segmentation and pronunciation effects. These findings will be explained with reference to factors defining the participants' learning environment, the transition of tertiary education to a competitive model, gender-specific religious and ethical affiliations, and the modified impression management hypothesis. The authors argue that a sufficient level of moral competence is critical for enabling prospective health providers to engage in demanding practical and normative contexts in healthcare facilities, where medical and clinical decisions are often interwoven with serious sociomoral aspects and responsibilities. A substantial number of medical solutions are still made on the basis of individual professional expertise and specialised skills on the one hand, *and*, on the other hand, through reference to individual moral judgments and decisions – even when collegial decision-making and centralized procedures try to relieve and support doctors and nurses as individual decision makers.

2. Moral Competence Definition

Making moral judgments and decisions has been intensively explored in philosophy, e.g., Aristotle's virtue, attempting to find the golden mean, universalizing an individual maxim "by virtue of the volition" (Kant 1785), prima facie choice (Audi 1999), judging "in accord with self-chosen principles" (Habermas), etc. Moral psychology addressed the ability to choose between 'right' and 'wrong' in a quick, intuitive or emotional modus (Haidt 2001, 2007) vs. a slow, deliberate, and explicit one. Finally, several noticeable approaches to the origins of moral competence were established, e.g., cultural-socialisational (Haidt 2001; Haidt & Kesebir 2010), developmental (Kohlberg 1964, 1984) and socio-evolutionary (Tomasello & Vaish 2013). Kohlberg, Blasi (1980), Lind (1978) and Rest (Rest 1986; Rest *et al*. 1999) began examining how the ability to make moral judgments manifests itself in subjects' judgmental behaviour. Four types of measuring instruments are available today (Ellemers *et al*. 2019, 337).

Lind's definition of moral competence focuses on a personal capacity to follow self-prioritized moral standards when making moral judgments or decisions, but definitely

1 The findings and hypotheses presented in this article were discussed at several congresses as peer-reviewed materials, especially at 14th International Symposium "Moral Competence: Its Nature, Relevance, and Education," Vilnius University, 23-24 July 2020, and the Annual Meeting of AERA "The Power and Possibilities for the Public Good When Researchers and Organizational Stakeholders Collaborate," San Francisco, 8-12 April 2021. We would like to thank Stephen Dersley (AMU Poznań) for his contribution to the linguistic shape of our research study.

breaks with linear or cumulative ('state-by-stage') development. Moral competence is a cognitive disposition pre-shaped by evolution; however, as moral competence varies from person to person, socio-educational factors are responsible for its improvement. Well-developed moral competence enables an individual to deal with sociomoral issues through making principled and deliberate judgments, without 1) an instant feeling of approval/disapproval, 2) a superficial 'yes' vs. 'no' opinion, 3) using violence, manipulation, or 4) conformity or submission to authoritarian others and powers.

According to Lind, moral competence has a dual-aspect structure. Making a moral judgment represents the first, cognitive aspect. In turn, moral orientations represent the second, affective aspect, which spurs a moral judgment maker to act accordingly and consistently. Both aspects of moral competence are distinct but highly correlated. Subjects differ little regarding their basic moral orientations, but greatly with regard to their moral competence. Studies have demonstrated that "the higher people's moral competence is, the more (…) they accept shared moral principles and reject low-type moral orientations" (Lind 2016, 62; see also Nowak 2016). High moral competence is a predictor of observing social rules and dealing with sociomoral controversies and conflicts. A morally competent individual remains open to the "unforced force" of better normative argument (Habermas in *Remarks to Discourse Ethics*), is willing to pursue agreement or to respect disagreement (for reasonable agreement to disagreement upon, e.g., the end-of-life decisions see [Wilkinson, Truog, & Savulescu 2016; Wilkinson & Savulescu 2018]). At this point, the following question arises: Do healthcare professionals use moral competence?

3. Moral Competence and a Paradigm Shift in Clinical Decision-Making

It is clear that standards of clinical decision-making evolve, and that "strong professions" (Helkama 2013) such as those involved in health care, do not always rely on an individual problem-solver's professional and normative expertise and competence (Helkama 2013, 99; Helkama, Uutela *et al.* 2003; Helkama & Ikonen-Varila 1996). The growing complexity of medical expertise and therapeutic evaluation has increased health providers' responsibilities. They might overwhelm individual capacities when it comes to dealing with current actionable decisions, some of them of cutting-edge type. To guide, justify and facilitate decision-making in complex clinical environments, procedural and relational ethic frameworks are increasingly established (e.g., Pollard 2015; deMartino 2017; Jenkins 2018). Bioethical boards, collegiality (Newton-Howes *et al.* 2019), asking "a senior colleague for advice" (Helkama 2013, 99), etc. or justification of a solution (Quenot *et al*. 2017). According to intersubjectivity, dialogue and discourse theories, in relational ethics there is no "I", only a "We": "We are the environment, 'we are the system'," Pollard (2015, 367) argues (see also Bergum 2005). For example, instead of caring or being responsible *for* a patient and deciding what is beneficial, right, etc., *for* her, decisions are made *with* patients (Pollard 2015, 362) and engage moral-discursive competences.

But a large number of clinical decisions, also sensitive and tough, are made by an individual health provider (deMartino *et al.* 2017; Deveterre 2010, 76; Jonsen 1995; Baumgarten 1980, 183), some of them being "coupled with the willingness to look at every new patient with fresh eyes and to view every illness as a unique case" (Löwy 1978, 130), some others come up with solutions in an emergency. Collective, procedural and technologically supported models and tools of clinical decision-making change, but doctors or nurses with their individual responsibilities remain constant and integral component of them. Both shared *and* individual decision-making requires moral competence. This competence shows the potential for overcoming exaggerated regulations, abstract principlism, hierarchies and power clusters characterizing health care facilities and health care as a public institution. Exaggerated regulations were criticized for "intrusion into the discretion of physicians about manner and extent of care, rationing of medical resources, and bureaucratic delays in meeting urgent needs" (Lee & Emmott 1978, 613; comp. also Haller & Stoelwinder 2017; Kashev 2016; Hills 2013; Allen & Harkind 2005; Freckelton & Petersen 2006; Hodges 2006; Hernandez & Medina 2005; Lind 2000a; Self & Baldwin 1994; Baumgarten 1980) and for suppressing subjects' ability to handle running challenges, risks, and uncertainty constructively (Pieniążek 2008, 129).

This may challenge decision makers' moral competence in clinical contexts. In 1895, the Polish physician Zygmunt Kramsztyk reported on such challenges:

> The more responsibility for other people that results from the profession one is engaged in, the more one is disturbed and one's mind occupied with the duties of that trade. The profession of physician is one of the most difficult from this point of view. (…) A feeling of discomfort which might spoil the night's rest, grounded in difficult, risky surgery; an irregular case of disease, constant suspense, mostly incomplete and seldom perfect results of treatment; unexpected complications, reproaches from patients, depressing albeit often unjustified remorse, unclear feelings of guilt: all of these are implanted in the everyday life of a physician and reflect his usual thoughts and feelings (Kramsztyk 1895/Löwy 1978, 146).

Health care related decision-making paradigms may shift, but the majority of clinical decisions in situ are made by human subjects by virtue of their individual skills and expertise. In last two decades the balance point in clinical and biomedical ethics shifted from abstract principles and procedures to making a concrete and situational decisions and arriving at solutions to problems in situ, by drawing on subjects' virtues abilities (Arthur *et al.* 2015; Kotzee, Ignatowicz & Thomas 2016; Kotzee & Ignatowicz 2016; Kaldijan 2014; Nucci & Narvaez 2008; Jansen 2000; Massingham 2019; Audi 1997). Subsequently, a revival of virtue ethics was proclaimed in academic biomedical ethics. These changes created a favourable background for justifying – and fostering – clinical and medical decision-making (especially in its socio-normative aspects) through moral competence: both in terms of educating subjects and conducting scientific research in developmental psychology. In this way, healthcare providers' (and future healthcare providers') moral competence deserves attention alongside virtue, and can be regarded

as a cognitive-affective concept that is supported by moral fortitude and virtue, as defined, e.g., in ancient Greek ethics. However, moral competence is not a plural phenomenon in the way that moral virtues are. There is only one moral competence, which is applicable in various sociomoral contexts, including healthcare professions. For instance, it is required to

- Help patients and care for them – as health providers represent helping professions – so "… a high level of moral judgment can be associated with seeing others' points of view and finding solutions instead of implementing the rules only" (Çiftçi & Yüksel 2010, 717),
- Deal with moral distress (e.g., institutional and political pressures, sociomoral climate) and negative social perception (Epstein & Whitehead 2019; Lamiani *et al.* 2017; O'Donnel *et al.* 2008),
- Manage decision-making shared with patients, their relatives, and further actors (Napiwodzka 2021; Entwistle *et al.* 2010),
- Deal with all types of sociomoral responsibilities and challenges produced in health professions, and
- Be "a good member of the medical profession" (Helkama 2013, 99) and team (Bate *et al.* 2012),

Below we present a pilot study with prospective health providers in Poland, followed by a discussion of the educational policies and factors responsible for their moral competence improvement.

4. Research Procedure

4.1 The Moral Competence Test (MCT)

The Moral Competence Test is an experimentally designed behavioral test to measure moral competence objectively and validly (Lind 1978/2020). In the first instance, the MCT sheds light on a participant's judgmental behavior. Because previous research had shown that rating arguments with regard to their moral quality – instead of with regard to whether the arguments match their opinions – is very difficult for most people, the MCT asks participants to rate (on a Likert scale from -4 to +4) arguments supporting and opposing the decision of a protagonist in two dilemma situations. Each situation triggers the feeling of a dilemma as a participant confronts two conflicting moral routes. Participants firstly rate a protagonist's decision. Subsequently, they rate six arguments which support this decision, and six which oppose it. Each argument represents one of the six types of moral orientation as defined by Kohlberg (1984). 24 arguments are independent variables in the MCT. For the standard MCT, the calculation of the C-scores only includes independent variables.

When participants show no moral competence, they indiscriminately accept all the arguments that support their own stance on the decision, and reject all opposing arguments. In contrast, when they have a high moral competence, they only accept arguments of high moral quality and reject those of law quality. They pay little attention to the question of whether the arguments match their opinions and reorient their attention toward principles and open their minds for agreement (or reasonable disagreement) with otherwise thinking persons.

In order to facilitate further analysis, each individual set of ratings (which literally visualizes a participant's moral behavior pattern) (Lind & Nowak 2015) is converted into a numerical C-score (C for competence), which ranges from 0 to 100. This conversion is made with the help of a multifactorial analysis of variance components. 0.0 designates no moral competence. Participants' C-scores vary widely but are mostly located in the lower part of the scale. C-scores above 40.0 are rare. Because consistent judgment is a function of a subject's moral competence and not a test property (and their judgmental behavior pattern can be regarded as their moral fingerprint), conventional criteria such as reliability and measurement error do not apply[2].

The MCT has been validated and certified for 40 languages. For its validation, four rigorous criteria are used: 1) In all the studies, people prefer the six types of moral orientations in the order which Kohlberg (1958, 1984) has predicted (see Fig. 2 in this paper). 2) These types also correlate with each other: neighbouring types are higher correlated than more distant types of moral orientations (which manifests itself as a quasi-simplex structure). 3) As mentioned above, the six type of moral orientation correlated very highly with moral competence, and 4) The MCT's C-scores cannot be faked upward, like the scores of moral preference or moral attitude tests (Lind 2002).

Theoretical and empirical validity have been repeatedly confirmed in numerous independent studies across countries and cultures. The MCT can be used repeatedly with the same participants if they are informed about the purpose. It is always used anonymously, so there is no reason to provide socially desirable answers. It can be used

2 The MCT was designed as an experiment with a multivariate orthogonal design, as E. Brunswik (1955) had suggested. This means that the MCT is not a "test" in the sense of testing psychology, but it is an *n=1 experiment* with three-factorial orthogonal design, operationalized as a questionnaire (Lind 1982). It has been designed to make the structure of moral judgments of individuals *manifest* and *visible*, and also to make it possible to quantify the degree to which the participants' moral orientations determine their responses. Because of this experimental design, the participants' pattern of responses let us directly *see* the properties of their moral competence without the aid of additional assumptions, as is the case with classical psychological tests. As a result, the criteria of classical tests do not apply. The participants' moral competence can develop and thus their scores can change. The MCT is never changed in any way. Thus, the MCT's equivalent to "reliability" is 1.0. The MCT's validity is checked using four well established psychological findings about the nature of moral judgment behavior: 1) The preference hierarchy of moral orientations, 2) The circumplex structure of their inter-correlations, 3) The correlational parallelism between moral competence on the one hand and the profile of moral orientations on the other, and 4) The non-fakeability of moral competence in experimental settings. These psychological (instead of purely formal) criteria provide more rigorous criteria for the validity of experimental designs than conventional statistical criteria used in test psychology. Rigorous means that the *a priori* probability of confirming these criteria by chance is extremely small, and, therefore, their confirmation is extremely informative (K. Popper).

with participants above the age of 10, especially with groups of learners, to evaluate the quality of trainings offered in educational settings.

4.2 Research procedure

The present pilot study is based on data collected in 2020 in Poland with the standard Moral Competence Test (MCT) which had been previously validated and certified for the Polish language. Approval was obtained from the Dean's Office at one of the medical universities. The MCT questionnaire was installed on the Survio.pl portal. Participants received the URL address via a bulk email managed by the department's officials. Participation remained voluntary, anonymized and randomized. Only adult participants were addressed. No personal or sensitive data were collected.

4.3 Participant sample description vs. feminization of health professions in Poland

A total final sample (size) of *n*=115 healthcare students were randomly surveyed, representing all years of study (1st to 6th year), their ages ranging from 19 to 45. Only Polish speaking students of full-time studies were addressed. 13,2% of respondents completed the MCT while 86,98% of the visitors did not. Male and female adults were addressed; however, with female participants *n*=88 and male participants *n*=27, a gender disparity-participation effect was noted.

The gender disparity effect can be explained by demographic tendencies recently observed in the medical workforce in Poland. According to Baliński and Krajewski (2018, 12), in 2016, for the total population of medical students (incl. healthcare students), 64% of female graduates and 36% of male graduates obtained their Licentia Medendi. The gender disparity effect observed with our sample corresponds with this tendency. The gender ratio of the total medical workforce population in Poland in 2016 was 0.58 (Baliński & Krajewski 2018, 41). Shannon *et al.* (2019) argue that feminisation trends in clinical and associated health occupations representing the traditionally male-dominated professions, in countries with under-resourced healthcare sectors including lower incomes, was triggered by emancipatory processes. This phenomenon was observed in Poland starting in the 2010s, when 80% of students of the medical university in which our pilot study was conducted were females. This tendency aroused a nation-wide discussion on introducing gender parities as one of the admission criteria for medical and healthcare students (Twardowska 2011). To this day no parities have been introduced.

In our study, the gender ratio of 76,5% may have affected data collection and data representativeness, due to the interest male and female subjects choosing to participate in the survey. For this reason, web surveys in general might be "susceptible to self-selection and reporting bias" (Shannon *et al*. 2019). Below we shall discuss the implications of the gender disparity for the research findings. Furthermore, voluntary participation of human participants in surveys may imply a positive selection effect.

Considering more detailed demographic characteristics of the medical university

involved in our study, it is worth noticing that in 2018, 255 students matriculated to their 1st year of study. In 2020, 317 studied all specializations offered at their 4th year of study (with an admission limit of 340 places for master studies) (Majewska 2020). At the same university, 164 students of the 6th year of study graduated.

5. Research Findings

In the following figures, the results of our study based on data collected from healthcare students (*N*=115) are displayed. Among all the years of study, years 2 and 3 seem most unfavourable for the development of students' moral competence:

Year of study	Mean C-score	N
1	32,89503	51
2	20,18794	2
3	18,14981	7
4	27,76710	29
5	33,05105	18
6	32,55845	8
All	30,48438	115

Figure 1: C-scores of the Polish healthcare students by year.

The C-scores of students of years 1, 4 and 5 were selected out in order to focus the study on the years represented by at least 18 participants:

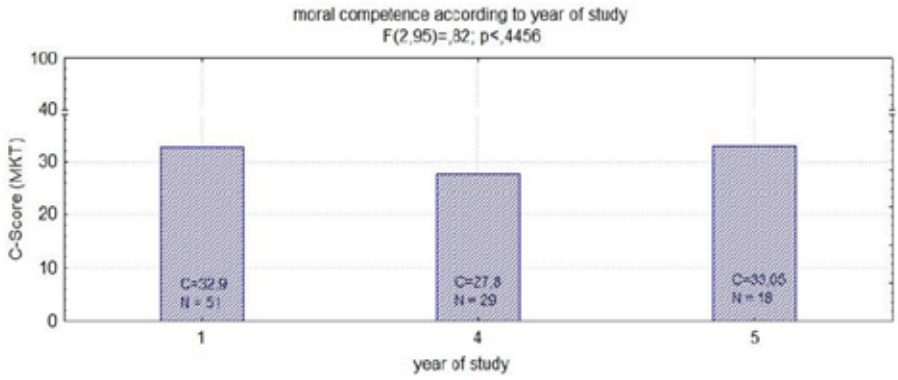


Figure 2: The C-scores of students of years 1, 4 and 5 represented by at least 18 participants were selected out.

Between the C-scores calculated for each individual dilemma separately, a difference was found: participants showed a slightly higher moral competence when rating arguments and counterarguments referring to the Workers' dilemma, while their C-score related to the Doctor's dilemma was lower (-3,7 points). A difference ≥ 8 points determines the segmentation effect. Unexpectedly, in the Polish sample, the segmentation effect was only slight:

	N	Main C-score	Minimum C-score	Maximum C-score	Standard deviation
Workers‘ dilemma C-score	115	49,53	1,81	92,35	22,32
Doctor‘s_dilemma C-score	115	47,66	0,00	92,55	25,22

Figure 3: C-scores of healthcare students calculated for each individual dilemma separately.

Further, another unexpected effect, namely a gender bias evident in female and male participants' C-scores (= 13,5 points) was found:

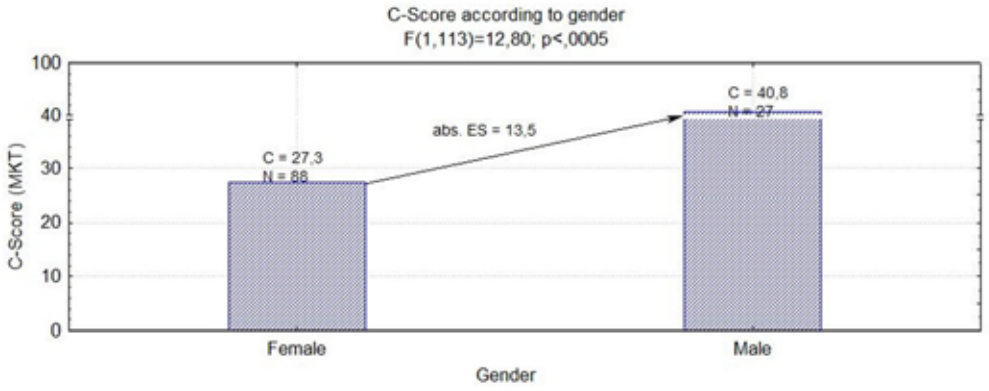


Figure 4: C-scores of healthcare students according to gender (a gender bias effect).

Finally, intergroup, gender-related biases manifested in C (for moral competence), scored separately for the Workers' and Doctor's dilemma, were identified in the sample. As shown in Fig. 5, the blue line demonstrates (1) a segmentation effect in female participants whose C-score for the Workers' dilemma is higher than for the Doctor's one, and (2) a pronunciation effect in male participants whose C-score for the Doctor's dilemma is higher than for the Workers' one:

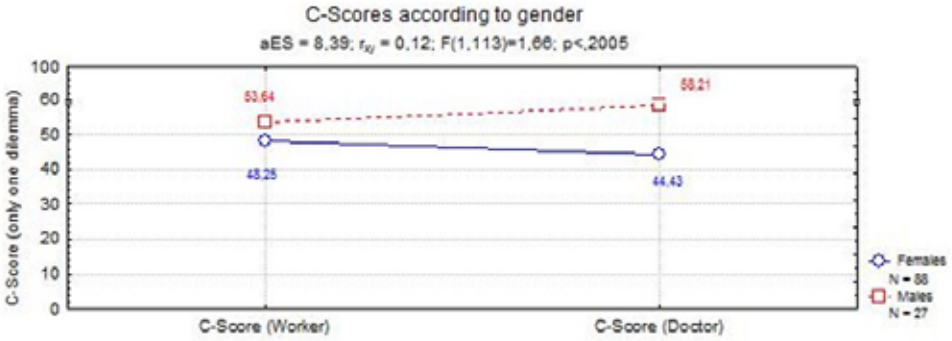


Figure 5: C-scores for Workers' and Doctor's dilemma by gender (segmentation and pronunciation effects), and the absolute effect size (AES) by gender for each individual dilemma: The Workers' dilemma C = 5,39 points; the Doctor's dilemma C = 13,78 points.

6. Explaining the Findings

6.1 The C-scores now and then

Due to the constant improvements made in Polish higher education, including in the medical and healthcare sciences, the researchers expected a change in the C-scores of healthcare students when compared with the former scoring (with MCT, at the same medical university, in the same faculty of healthcare) conducted in 2007/8 as a part of MCT validation study for the Polish language (Nowak et al. 2007/8, total N=370). In the related study, healthcare students were represented by N=112 and a slight decrease of their C (also regarded as a stagnation effect) between the 1st and 9th study semester was observed:

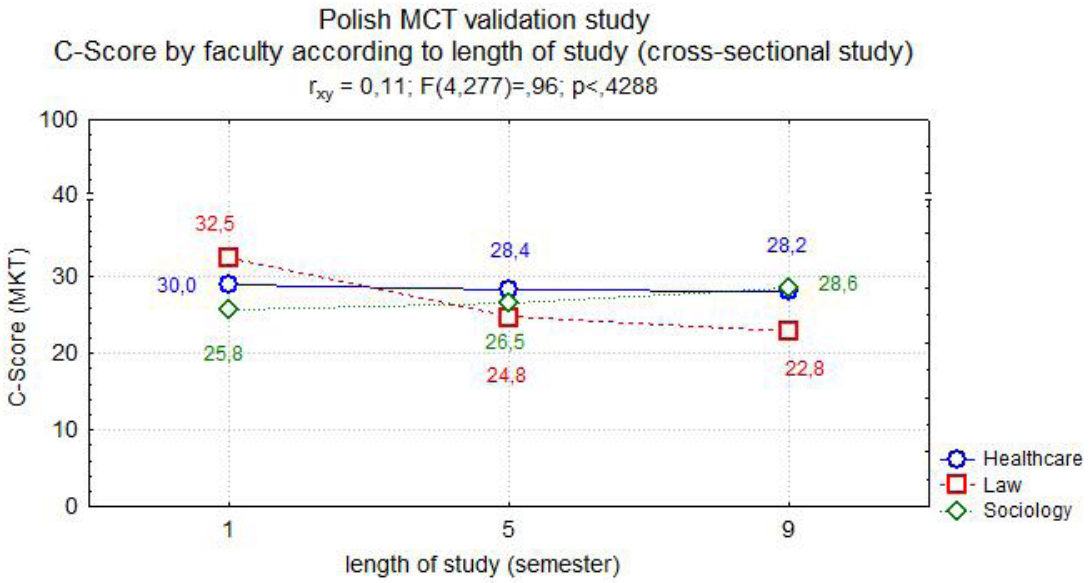


Figure 6: Students' moral competence by faculty and by semester, scored in 2007/8 in Poland.

Comparing the initial scores (1st year of study) in 2007/8 (Fig. 6) and 2020 (Fig. 1 and 2), an increase of C = 2,9 (2020: C = 32,9; 2007/8: C = 30,0) can be noticed. This increase does credit the high education quality as fostering moral competence of students stronger than in 2007/8. However, the moral competence developmental trajectory in 2007/8 was more linear than that identified in 2020, as no dramatic decrease between the 1st and 5th semesters of study was observed.

Furthermore, the final scores measured for the 9th semester of study in 2020 (C = 33,05) and 2007/8 (C = 28,2) show an increase of 4,85 points. C-scores can be considered in a more international context:

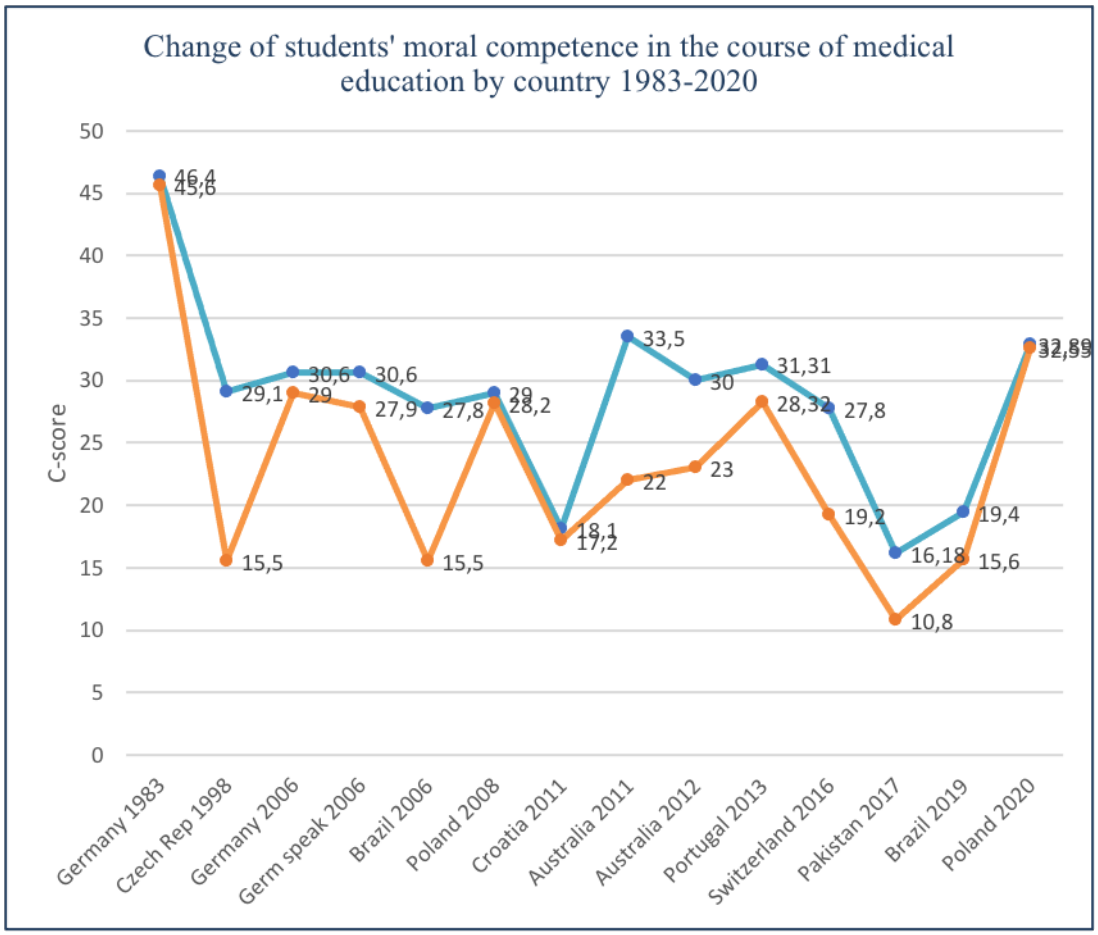


Figure 7: Changes in the moral competence of medical and healthcare students by cohorts representing 9 different countries (Sources: for Germany 1983: Lind 2000; for The Czech Republic 1998: Slovàckovà 1998/9; for Germany 2006, German speaking Switzerland 2006 and Brazil 2006: Schillinger 2006; for Poland 2007/8: Nowak et al. 2007/8, MCT validation study published in 2013; for Croatia 2011: Kukolja Taradi 2011, a supervised MCT validation study; for Australia 2011 & 2012: Hegazi and Wilson 2013; for Portugal 2013: Feitosa et al. 2013a; for Switzerland 2016: Hummel et al. 2016; for Pakistan: Abassi et al. 2017; for Brazil 2019: Castro 2019; for Poland 2020: as reported in this article).

A set-off between each two vertically linked values (indicated on the blue and orange curves) demonstrates how medical students' moral competence has changed between the initial and the final C scored in medical schools representing international contexts between 1983 and 2020. In both the Polish studies, the set-off was slight and the C-scores were comparatively high. This set-offs depicted in Fig. 8 suggest two dominant tendencies: (1) Stagnation, and (2) A small to high decline of students' moral competence during their medical education. The tendencies show persistence across a variety of research designs, cohorts' characteristics, educational cultures and policies, training methods, socio-economic and demographic backgrounds, etc. Results showing a change of $\geqslant 5$ C-points can be regarded as significant. Several of the total of fourteen studies show a significant decrease of moral competence as a possible result of medical education, while the majority of studies shows stagnation or decrease. Because these tendencies have been found in cohorts representing medical universities in such different educational cultures over different periods of time (as reported, e.g., by Rego

& Bataglia 2017; Slovàckovà & Slovàcek 2007; Helkama *et al.* 2003; Lind 2000a, 2000c; Lind & Schillinger 2003; Rego *et al.* 2011a, 2011b), it seems that the failure of medical and healthcare education to improve moral competence is a more frequent phenomenon, though the authors of this study would not generalize its omnipresence. The MCT studies with students representing other fields of higher education show an increase or stability in moral competence (Lind 2002; Nowak 2013).

Subsequently, we attempt to explain the changes in the Polish medical students' moral competence, i.e., the alternating effect of regress vs. progress (Fig. 1 and Fig. 2) as a multifactorial analysis of the learning environment surrounding the participants of our study.

6.2 A multifactorial analysis of learning environment to explain the C-scores obtained with the pilot study

As the C-scores of healthcare students measured 2020 in Poland show an unusual developmental dynamic, we analyse the factors possibly having an effect on this dynamic. As already suggested, we shall skip changes between the underrepresented years of study, and focus on the C-scores of years 4 and 5:

year 4: C = 27,76; year 5: C = 33,05

(and year 6: C = 32,55; here only as a trend value).

Years 1, 2 and 3 follow the pre-clinical educational curriculum, whereas years 4, 5 and 6 are on the clinical curriculum. In all years (1 to 6) students are involved in practice (120 hrs per year). Their normative education includes: Ethics (10 hrs, 1st year), Clinical Procedures and Professionalism (25 hrs, year 1), Medical Law and Forensic Medicine (10 hrs in year 3; 50 hrs in year 5).

Year 3 includes Medical Simulation with Standarized and Simulated Patient training (10 hrs) which continues during year 4 (20 hrs) and year 5 (25 hrs). This training programs engage volunteering actors trained by professionals to act out symptoms and engage in dialogue with healthcare students (source: https://csm.ump.edu.pl/aktualnosci/operacja-symulacja-1; https://www.ump.edu.pl/komunikat/projekt-symulowany-pacjent). This educative innovation was implemented in 2016. We suggest that it created a favourable and safe learning opportunity to promote participants' moral competence between the 4th and 5th year of study. Participants who reported on their experiences with PBL and simulated patient training showed higher C-scores than those without such experience. A difference of 6,4 points was stated:

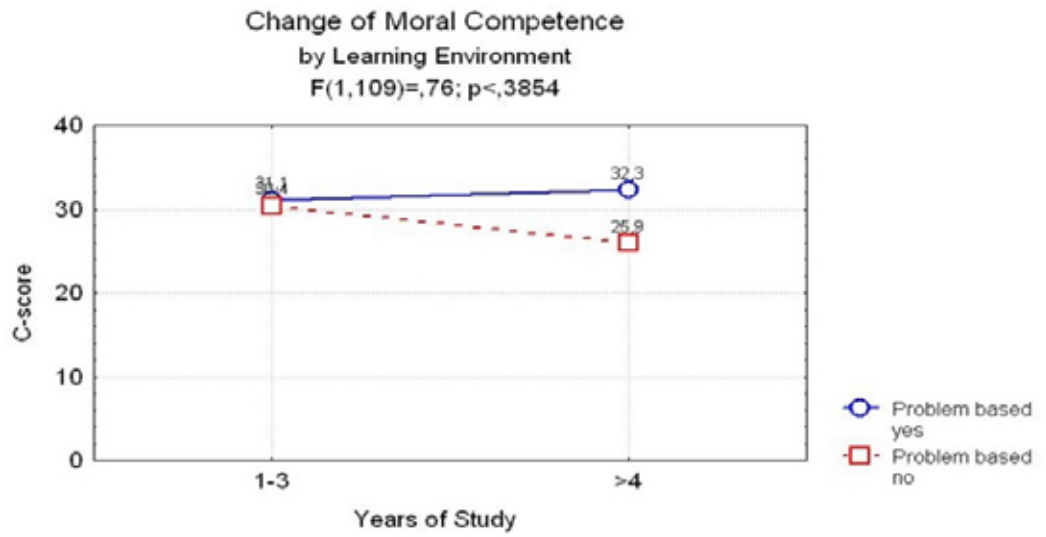


Figure 8: The effect of problem-based and simulated patient training on healthcare students' moral competence, measured in 2020 in Poland.

Although PGL efficiency was assessed as the "empty glass" effect (Gomes & Rego 2011, 561; Orsolya, Hemmerling, *et al.* 2017; Shanley 2007; Shamsan & Syed 2009; Feitosa *et al.* 2013b), combining PBL with such learning methods as dealing with health-related issues situated in contexts which include a simulated patient, role-taking, responsibility-taking and guided reflection seem to improve learners' moral competence. Such a combination shows higher efficiency than academic ethics courses (e.g., Meireles Martins *et al.* 2020; Rzymska *et al.* 2014; Hegazi & Wilson 2013; Campbell & Chin 2007; Langer *et al.* 2015).

Scholars stress that years 4 to 6 are generally demanding for medical and healthcare students. They experience crisis while crossing a "professional Rubicon" (Sandor *et al.* 2015; Abassi *et al.* 2017, 137) and confront contexts challenging their professional expertise, sociomoral and personal competencies. As a result, a loss of ideals, cynicism, egocentrism (Self *et al.* 1993), competitive strategies – and the moral competence decrease can be observed (Serodio *et al.* 2016; Schillinger 2006; Lerkiatbundit *et al.* 2006; Hodges 2006; Hernandez & Medina 2005). Education should offer remedies against such a crisis, to strengthen students and their moral competence during this critically important career stage (Lind 2015, 2002, 2000a, 2000b, 2000c).

6.3 Tertiary education on transition to competitive model (with implications for medical and healthcare studies)

Competitiveness is closely related to the selection of candidates, as those with the highest test scores recalculated to credits obtain their admission passports. Competitiveness in medical education is a more general phenomenon, with serious consequences for moral competence development. To achieve as high a ranking as possible, accreditations, parametrizations, etc., universities must constantly focus on the official academic indicators (and the strategies and policies boosting them), while the

C-score remains unofficial and low-prioritized. The same applies to trainings to foster students' moral competence, as the latter does not count in evaluations. "What is called 'quality' could also be understood as 'measures of institutional advantage'," as Schillinger (2006, 115) explains with reference to the sample of Brazilian students representing Higher education, including the medical sciences, was for decades discussed in terms of increasing competitiveness. In the last two decades, the tertiary education sector in Poland faces radical modernization and internationalization, corporatisation and commercialization to be "much more strongly linked to the labour market" (Kwiek 2012, 349).

As a result, higher education aspires to implement more competitive trainings and programs. The continuity across all these radical changes is the competitiveness of medical studies for candidates. "Social competition for the most valuable student places in most prestigious institutions is clearly increasing. However, as elsewhere in HPS, the intensity of that competition (which occurs in full-time taxation-financed studies in the public sector only) is highest in the traditionally least accessible faculties of law and, outside of comprehensive universities, the faculties of medicine in specialist universities. For instance, in 2016 there were on average 16.8 candidates per vacancy in medical studies" (Kwiek 2018, 351). Advocated by scholars and state policies, this kind of social competition rather resembles social segregation than "fair education" and opportunity equality according to, e.g., Giesinger (2011, 2009).

Competitiveness is closely related to the selection of candidates, as those with the highest test scores recalculated to credits obtain their admission passports. Competitiveness in medical education is a more general phenomenon, with serious consequences for moral competence development. To achieve as high a ranking as possible, accreditations, parametrizations, etc., universities must constantly focus on the official academic indicators (and the strategies and policies boosting them), while the C-score do not count in evaluations. "What is called 'quality' could also be understood as 'measures of institutional advantage'," as Schillinger (2006, 115) points out with reference to the Brazilian C-scores representing low- vs. high-competitive medical schools.

Scholars posit that moral competence regression observed with medical and healthcare students to be caused by the increasing competitiveness and corporatization of the style of medical education, in both the public and private education forms represented by medical universities (e.g., Rego 2004; Rego *et al.* 2011a; Rego *et al*. 2011b; Ladim *et al.* 2015; Feitosa *et al.* 2013a, 2013b; Pascarella 1991; Pascarella & Terenzini 1991). Another type of the "competitive culture" can be observed in medical education, whose effects were evaluated as "destructive for students" (Yocom 2018) and compared to social Darwinism (Kohn 1986). According to Kohn, medical schools prioritize "competition over cooperation" which intensifies, for instance, "anxiety, self-servitude and weak interpersonal skills," "whereas cooperation fosters a better learning environment" (Yocom 2018) and makes it more favorable for moral competence development and further core skills. Originally

identified in social and educational contexts in the USA (Kohn 1986), competitiveness infiltrated the policies and practices in higher education within the European Union in the 2010s. In the 2020s, as former medical students in Poland report, once students were admitted to their field of study, they exclusively focus on not failing and losing their place (Zdziebko 2020). This would be one of explanations for the decreasing moral competence in years 2 and 3 (in year 1 students present the developmental effects gained in their high schools).

Polish tertiary education followed these trends only in the last decade, under different economic, systemic and sociocultural circumstances, e.g., with a medical education system that was mainly public (the first two non-public medical universities were established in 2017 and 2018), and with a shortage of medical staff in public healthcare vocations. Structural and psychosocial factors increasing the high level of competition between medical students (as described by Yocom) are documented in a few scholarly sources. E.g., in 2012, 9% of students at the Medical University in Wroclaw (Poland) declared their professional choice as motivated by economic interests; 21% of them had chosen the medical professions due to their promising perspectives toward employment and career (Waszkiewicz *et al.* 2012; compare Matyja *et al.* 2012). Hryniewicz's (2016) trial with medical students (*N*=206) showed that economic motivations outweighed all other incentives, e.g., training for a caring profession, finding purpose or passion, etc. Still, discussing statistical facts, structural and societal tendencies are not sufficient to explain and predict how medical students' moral competence would change in the course of their education. A more detailed, multifactorial analysis of a distinct learning environment will follow in the next subsection.

6.4 Moral competence and gender-specific challenges

The gender effect on moral competence is described as ambiguous and disputable (e.g., Kohsravi Zadanbeh & Zakerian 2011). Rest (1986) and Schillinger (2006) have found that females' C-scores are higher than males', although "moral judgment competence levels are similar for men and women (MJT C-score = 24 and 23, respectively). Results from a one-factor analysis of variance and effect size are: F(1,1142) = 2,109, p = 0,15, r = 0,04. According to Schillinger, no significant gender differences were found regarding subject's preferences for the moral stages either (affective aspect of moral behavior)" (Schillinger 2006, 98).

In our sample, a significantly different effect was uncovered. As demonstrated in Figure 4, the moral competence index of male participants C = 40,8 points (*N*=27), yet C = 27,3 points was measured for female participants (*N*=88); the absolute difference = 13,5 C-points.

We do not see reasons for the biological gender reference (Becker & Ulstad 2007; compare Singh *et al.* 2002) to explain this phenomenon. Furthermore, we agree with Tirri and Nokelainen's (2007) conclusion that "gifted students in science may have the best

cognitive skills and logical thinking but they may lack the ethical sensitivity that is needed to solve moral dilemmas in science" (Tirri *et al.* 2012, 6) and social contexts.

Following Becker's and Ulstad's hypothesis, we suggest the following explanation: the gender gap between the male and female minority's C-scores can be explained by females "engaging in impression management with their survey answers" (Becker & Ulstad 2007, 88). "There is a pervasive tendency to present oneself in the most favorable light relative to prevailing social norms," or – in our opinion – to present oneself according to the prevailing competitive strategies. "This interest in answering in a socially desirable manner is known as impression management" (Becker & Ulstad 2007, 78). Thus impression management can be one of the competitive tactics employed by the female participants of our study.

Becker's and Ulstad's hypothesis can be strengthened by additional interviews related to the investigated academic context. In their childhood many females in Poland were reportedly rewarded for being polite, humble, submissive, caring, and trying to gain others' recognition (Zdziebko 2020). Secondly, female students who were additionally interviewed for ethnographic background of the surveyed group, reported that female medical students employ tactics to make survey results more impressive, and to boost self-promotion and self-identification regardless of interviewing methods used (e.g., paper-and-pen surveys vs online surveys; open vs closed questions). When applied to the Moral Competence Test, such strategies imply a countereffect, i.e., low C index.

As Hren *et al.* demonstrated, Croatian female students scored with DIT and instruments measuring Machiavellian tactics on socially desirable responding behaviors "had higher scores on the SDR impression management subscale, whereas male students scored higher on the self-deception subscale; a finding similar to other studies on college students. Impression management is a construct of a deliberate attempt to present a socially favorable personality, whereas self-deception is an overly positive but honest bias in self-description" (Hren *et al.* 2006, 274). Furthermore, evidence for medical students' verbal strategies to "impress senior medical staff, which was directly seen to prepare the way for prestigious jobs in the future. More subtly, some students used phrases during the interviews which implied some advantage over other students" was found by Lempp and Seale (2004, 772). As yet, research like this has not been conducted in Poland, but the authors suggest considering such phenomena to explain low C-scores of surveyed female participants.

6.4.1 A segmentation effect and gender

Wakenhut (1982), Lind (2000d), Schillinger (2006), Hegazi and Wilson (2013), Bataglia & Schillinger (2013), Feitosa *et al.* (2013a; 2013b) observed a segmentation effect between C- values scored for each dilemma separately. In fact, the Doctor's dilemma can be more challenging for participants than the Workers' one. However, segmentation is not

always related to low moral competence. Rather, situational, institutional, socialisational and individualistic factors (also interpenetrating) enter into the equation, e.g., "different interpretations of the situation, closeness to everyday-life, personal experience, gender role, institutional pressure, and ideological barriers to autonomous judgment" (Lind 2000d, 1). Furthermore, a person's moral thinking might be controlled by institutional or ideological authorities, or narrowed to a specific professional ethics such as medical ethics (with regard to abortion, euthanasia, in vitro fertilisation, etc.). As a result, persons facing one of the dilemmas can be afraid of their independent reflection and judgment, or their minds turn to quasi-Freudian superego censorship.

Previous research shows that "religiously oriented subjects suppress their autonomous moral judgment on dilemma contents, on which the church takes a strong stance" (Lind 2000, 4) and submit adequate assessments concerning all pro-arguments to their 'disapproval'. Hegazi and Wilson (2013) argue that a segmentation effect with respect to the Doctor's dilemma caused medical students' moral competence to drop at one of the Australian universities. However, participants with higher C-scores were able to show more resilience to the segmented judgment.

In terms of the MCT methodology, a segmentation effect can be identified when participants' C-score is lower for the Doctor's dilemma (at least 8 C-points less) than for the Workers' dilemma. When the C-score for the Worker's dilemma is lower than for the Doctor's dilemma, a pronunciation effect can be noticed. In this study, the segmentation effect of C = -3,7 for the entire surveyed sample remains minimal and non-significant (lower than 8 C-points). However, as depicted in Fig. 5, moral competence separately scored by gender and then by each individual dilemma, is distinguished by (1) a segmentation effect in female participants and by (2) a pronunciation effect in male participants.

In females, a segmentation effect was = 5,1 C-points and higher than for the entire sample, but still non-significant. It was just demonstrating that female healthcare students were more challenged by the Doctor's dilemma than by the Workers' one.

However, a reverse pronunciation effect was observed in the male healthcare students participating in the study. They dealt better with the Doctor's dilemma regardless of the fact that the latter is more demanding than the Workers' one.

Eventually, the final absolute effect size (AES) as the final C-score gained in the course of studies) – only for the Doctor's dilemma – was identified as significantly higher for males (C = 13,78) than for females. The AES for the Workers' dilemma was C = 5,39.

This gender-related biases can be explained by various hypotheses. For example, M. Wnuk (2010; see also Parchomiuk & Byra 2015) examined medical female students' spiritual and religious strategies for coping with stress and moral distress in Poland. The mean scores for praying frequency (by gender) were 3.28 (for females, *N*=315, standard deviation 1.34) and 2.75 (for males, *N*=50, standard deviation 1.44) (Wnuk 2013). This

may result from euthanasia-related controversies not being equally demanding for female and male healthcare students, so they can develop different moral strategies to cope with them. Wnuk's findings can be confirmed by Pew Research Center's global study, which shows that females are generally more likely to affiliate with religious beliefs than males (see Pew Research Center Religion & Public Life 2016); gender disparity of this type also applies to Roman Catholicism.

As Capraro and Sippel (2017) argue, certain types of dilemmas imply a strong emotional response in accordance with a gender-specific ethical affiliations, e.g., socially responsive ethics (care, empathy, compassion, altruism) vs. principled ethics that may abstract from interindividual and relational social contexts). The Roman-catholic affiliation still present in the cultural background of the Polish sample would be more in line with responsive ethics, which is still predominant in Polish healthcare universities and is increasingly opposed to ethics based on principles that go beyond particular confessions and involving, e.g., human rights and universally accepted principles. If principles are conflicting in a subject's moral mind, this may seriously challenge their moral competence, firmness, and emotionality. In such cases the MCT participants would tend to strongly disagree even with high-type normative arguments, and face confusion and helplessness as moral decision makers.

A doctor's dilemma contained in the MCT can be demanding in a twofold context: (1) religious and (2) medical. These contexts can be regarded as connected in cultures with more traditional ethical and religious affiliations. But they can be regarded as disconnected in secular and secularizing contexts, too, where "saving human life at all costs and to its very end" (Ostrowska 1991, 65) remains the only admissible way of conduct despite principle plurality (some principles making the objection against futile therapy well justifiable).

In Poland, medical students' beliefs and attitudes concerning this topic could uncover some interesting changes (Lurka 2020; Szadkowska-Szlachetka *et al.* 2019). Between 2010-2020 medical students' acceptance of euthanasia oscillated between 45% – 51.3% (according to our data). However, it is not just beliefs and attitudes, but the ability to judge in line with high-quality normative criteria chosen to justify one's judgment and do so beyond attitudes, beliefs, and otherwise 'initial impulses', which instantly 'breaking into pieces', as D. Hume put it in *A Treatise of Human Nature*. It is precisely this that contributes to moral competence and is scored with the MCT. Furthermore, it is also respect for opinion, value and principle-pluralism. Advocating for a reasonable dissensus with regard to end-of-life decisions, Wilkinson *et al.* stress how crucial it can be for physicians (and all subjects in general) "to fairly appraise the options" and "different conclusions," and "to understand the nature of disagreement" (Wilkinson *et al.* 2016, 118), especially when the two conflicting options making up a dilemma can be justified by equally high-quality moral principles. Philosophy defines the justification and acceptance of a reasonable disagreement as an achievement of deliberative and democratic competence. According

to our long-term expertise in education based on Lind's dilemma discussion plan (the Konstanz Method of Dilemma Discussion), we may confirm that participants with higher moral competence deal much better with normative dissonance and are more likely accept and justify disagreement with high-level normative orientations which result in reasonable and principled disagreement.

Conclusions

The main objective of this article was to report on, and to examine changes in, the moral competence in healthcare students (*N*=115) representing a distinct learning environment in Poland. Being in transition to a competitive tertiary education model, this environment seems to incorporate a wide range of ambiguous factors that affect students' moral competence in various ways. During their pre-clinical curriculum, a decrease of "C" (for moral competence) was noticed. The clinical curriculum had a beneficial effect on students' moral competence. However, female participants manifested lower moral competence while dealing with the end-of-life dilemma than males. This phenomenon was identified as a gender bias, as in previous studies females often showed higher C-scores than males. Further, a segmentation effect and a pronunciation effect were identified and discussed on the basis of religious and non-religious ethical affiliations and principle pluralism.

Also, a modified impression management hypothesis was introduced as a gender-specific, competitive strategy to explain the overall low C-scores of female participants.

On the basis of a thorough, up-to-date literature review, healthcare students' moral competence improvement was documented as underprioritized for competitive medical education in several countries. The authors consequently argued for educating moral competence as a skill of critical importance for decision makers who deal with clinical-moral decisions that are made regardless of paradigm shifts.

References

Abassi W. G., Shakir S., *et al.* 2017. "Effect of Medical Education on the Moral Competence of Medical Students," *Proceeding of the 4th Int Conference on Arts and Humanities* 4:127–140. DOI: 10.17501/icoah.2017.4113.

Adams V., Behague D., *et al.* 2019. "Re-imaging Global Health through Social Medicine," *Global Public Health* 14(10):1383–1400. DOI: 10.1080/17441692.2019.1587639.

Agurto M., Tello D., *et al*. 2017. "Índice C en médicos de dos centros hospitalarios chilenos según el test de competencia moral de Lind," *Revista Médica de Chile* 145:1122–1128.

Allen D. & Harkind K. J. 2005. „Too Much Guidance?" *The Lancet* 365(9463), #P1768, May 21.

Andrews T. 2006. "Conflicting Public Health Discourses – Tensions and Dilemmas in Practice: The Case of the Norwegian Mother and Child Health Service," *Critical Public Health* 16(3):191–204.

Arthur J., Earl S. R., Thompson A. P., *et al.* 2021. "The Value of Character-Based Judgment in the Professional Domain," *Journal of Business Ethics* 169:293–308. DOI: 10.1007/s10551-019-04269-7.

Arthur J., Kristjànsson K., Thomas H., *et al.* 2015. „Virtuous Medical Practice: Research Report." Foreword by M. Pringle. Birmingham: Birmingham University, The Jubilee Centre for Character and Virtues.

Audi R. 2018. "Understanding, Self-evidence, and Justification," *Philosophy and Phenomenological Research* 99(2):358–381.

Audi R. 1997. *Moral Knowledge and Ethical Character.* Oxford: Oxford University Press.

Baliński P. & Krajewski R. 2018. *Lekarze i dentyści w Polsce – charakterystyka demograficzna. Stan w dn. 31.12.2016 [Healthcarers and Dentists in Poland – a Demographic Approach. The State of the Art on 31 Dec 2016].* Warszawa: NIL. URL: https://nil.org.pl/uploaded_images/1575630176_demografia-2016.pdf (retrieved on Jan 21, 2021).

Bataglia Unger P. & Schillinger M. 2013. "Moral Segmentation in Studies with the Moral Judgment Test in Brazil," in E. Nowak, D. E. Schrader, B. Zizek (Eds), *Educating Competencies for Democracy* (pp. 71–82). Frankfurt am Main – Bern – New York: Peter Lang Editions.

Bate L., Hutchinson A., Maskrey N. 2012. "How Clinical Decisions Are Made," *British Journal of Clinical Pharmacology* 74(4):614–620.

Baumgarten E. 1980." The Concept of 'Competence' in Medical Ethics," *Journal of Medical Ethics* 6:180–184.

Becker D. A. & Ulstad I. 2007. „Gender Differences in Student Ethics: Are Females Really More Ethical?," *Plagiary: Cross-disciplinary Studies in Plagiarism. Fabrication, and Falsification,* 77–91.

Bergum V. & Dossetor J. 2005. *Relational Ethics. The Full Meaning of Respect*. Hagerstown, MD: University Publishing Group.

Bergum V. 1992. „Beyond Rights: The Ethical Change," *Phenomenology + Pedagogy* 10: 75–84.

Blasi A. 1980. „Bridging Moral Cognition and Moral Action: A Critical Review of the Literature," *Psychological Bulletin* 88:1–45.

Braddock C. H. & Snyder L. 2005. "The Doctor Will See You Shortly," *General Internal Medicine* 20(11):1057–1062.

Brunswik E. 1955. „Representative Design and Probabilistic Theory in a Functional Psychology," *Psychological Review* 62(3):193–217.

Campbell A. V., Chin J., *et al.* 2007. "How Can We Know That Ethics Education Produces Ethical Doctors?," *Medical Teacher* 29(5):431–436. DOI: 1080/01421590701504077.

Capraro V. & Sippel J. 2017. "Gender Differences in Moral Judgment and the Evaluation of Gender-specified Moral Agents," *Cognitive Process* 9. DOI: 10.1007/s10339-017-0822-9.

Carver R. P. 1993. "The Case Against Statistical Significance Testing Revisited," *Journal of Experimental Education* 61(4):287–292.

Castro M. R. 2019. "Avaliação da competência moral de estudantes de Medicina." Dissertação Belo Horizonte, Faculdade de Medicina, University of José do Rosário Vellano (data and findings peer-reviewed).

Çiftçi N. & Yüksel A. 2010. "Analysis of the Relationship between Moral Judgment Competence and Empathic Skills of University Students," *Educational Sciences: Theory & Practice* 10(2):707–724.

Coggon J. & Miola J. 2011. "Autonomy, Liberty, and Medical Decision-making," *The Cambridge Law Journal* 70(3):523–547.

Conroy M., Malik Y. A., Hale C., *et al.* 2021. "Using Practical Wisdom to Facilitate Ethical Decision-Making: A Major Empirical Study of Phronesis in the Decision Marratives of Doctors," *BMC Medical Ethics* 22, #16.

Corell A., Regueras L., *et al*. 2018. "Effects of Competitive Learning Tools on Medical Students: A Case Study," *PlosOne*, March 8. DOI: 10.1371/journal.pone.0194096.

DeMartino E. S., Dudzinski D. M., & Kramer D. B. 2017. „Who Decides When a Patient Can't? Statutes on Alternative Decision Makers," *The New England Journal of Medicine* 376(15):1478–1482.

Deveterre R. J. 2010. "Practical Decision Making in Health Care Ethics: Cases and Concepts." Washington, DC: Georgetown University Press.

Dochy F., Segers M. *et al.* 2003. „Effects of Problem-based Learning: A Meta-analysis," *Learning Instruction* 13(5):533–568.

DuBois J. 1997. "The Moral Judgment of Medical Personnel. A Cross-cultural Study about Brain Death and Organ Explantation." Unpublished dissertation, University of Vienna.

Dwarak I. *et al*. 2013. "Spirituality and Dogmatism: Social Implications of Dogmatism and Its Cure," *Journal of Humanities and Social Science* 16(6):86–94.

Ellemers N., van den Toorn J., & Paunov Y. 2019. "The Psychology of Morality: A Review and Analysis of Empirical Studies Published from 1940 through 2017," *Personality and Social Psychology Review* 23(4):332–366.

Entwistle V. A., Carter S. M., Cribb A., & McCaffery K. 2010. "Supporting Patient Autonomy: The Importance of Clinician-patient Relationships," *Journal of General Internal Medicine* 25(7):741–745. DOI:10.1007/s11606-010-1292-2

Epstein E. G., Whitehead P. B., *et al.* 2019. "Enhancing Understanding of Moral Distress: The Measure of Moral Distress for Health Care Professionals," *AJOB Empirical Bioethics* 10(2):3111–3124. DOI: 10.1080/23294515.2019.1586008

Feitosa H., Rego S., *et al.* 2013a. „Competência de juízo moral dos estudantes de medicina: um estudo piloto" [Moral judgment competence of medical students: a pilot study]. *Revista Brasileira de Educação Médica* 37(1) (no pagination). DOI: 10.1590/S0100-55022013000100002.

Feitosa H., Rego S., Bataglia P. U., *et al.* 2013b. „Moral Judgment Competence of Medical Students: A Transcultural Study," *Advances in Health Sciences Education* 18(5):1067–1085. DOI: 10.1007/s10459-013-9449-5

Freckelton I. & Petersen K. 2006. *Disputes and Dilemmas in Health Law.* Sydney: The Federation Press.

Gardiner P. 2003. "A Virtue Ethics Approach to Moral Dilemmas in Medicine," *Journal of Medical Ethics* 29:297–302.

Giesinger J. 2011. "Education, Fair Competition and Concern for the Worst Off," *Educational Theory* 61(1):41–54.

Giesinger J. 2009. "Freie Schulwahl und Bildungsgerechtigkeit," *Zeitschrift für Erziehungswissenschaft* 12:170–187.

Gilligan C. 1977. „In a Different Voice: Women's Conception of the Self and Morality," *Harvard Education Review* 47:481–517.

Gilligan C. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Gilligan C. 1986. „*On In a Different Voice*: An Interdisciplinary Forum: Reply," *Signs* 11:324–333.

Gomes A. P. & Rego S. 2011. « Transformação da educação médica: é possível formar um novo médico a partir de mudanças no método de ensino-aprendizagem?» [Transformation of medical education: can a new physician be trained based on changes in the teaching-learning method?], *Revista Brasileira de Educação Médica* 35(4):557–566. DOI: 10.1590/S0100-55022011000400016.

Habermas J. 1998. "Reply to Symposium Participants," in M. Rosenfeld & A. Arato (Eds), *Habermas on Law and Democracy* (pp. 381–452). Berkeley, CA: University of California Press.

Habermas J. 2015 (6th Edition). *Erläuterungen zur Diskursethik*. Frankfurt am Main: Suhrkamp.

Haidt J. 2007. „The New Synthesis in Moral Psychology," *Science* 316:998–1002.

Haidt J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108:814–834.

Haidt J. & Kesebir S. 2010. „Morality," in S. Fiske, D. Gilbert, & G. Lindzey (Eds), *Handbook of Social Psychology* (pp. 797–832). Hoboken, NJ: Wiley.

Haller G. & Stoelwinder J. 2017. "Is 'Crew Resource Management' an Alternative to Procedure-based Strategies to Improve Patient Safety in Hospitals?," in M. Bourrier & C. Bieder (Eds), *Trapping Safety into Rules. How Desirable or Avoidable Is Proceduralization?* London: CRC Press. DOI: 10.1201/9781315549774

Hegazi I. & Wilson I. 2013. "Medical Education and Moral Segmentation in Medical Students," *Medical Education* 47:1022–1028. DOI: 101111medu.12225

Helkama K. 2013. "Change in Moral Judgment in Medical School: The Role of Hierarchy," in E. Nowak, D. E. Schrader, & B. Zizek (Eds), *Educating Competencies for Democracy* (pp. 97–102). Frankfurt am Main – Bern – New York: Peter Lang Editions.

Helkama K., Uutela A., Pohjanheimo E., *et al.* 2003. „Moral Reasoning and Values in Medical School: A Longitudinal Study in Finland," *Scandinavian Journal of Educational Research* 47(4):399–411.

Helkama K. & Ikonen-Varila M. 1996. "Moral Stages and Principles of Justice in Real-life Moral Decision-making Situations," poster presented at the 14th Biennial ISSBD Conference, Quebec City, 12-16 August.

Hernandez C. & Medina G. 2005. "Ethics and Professionalism in Medicine: Cross-curricular Integration of Ethical Basis for Medicine Students," presentation at the AME Annual Meeting in Cambridge, MA, Nov 2005.

Hodges B. 2006. "Medical Education and the Maintenance of Incompetence," *Medical Teacher* 28(8):690–696.

Hren D., Vujaklija A., *et al.* 2006. "Students' Moral Reasoning, Machiavellianism and Socially Desirable Responding: Implications for Teaching Ethics and Research Integrity," *Medical Education* 40:269–277.

Hummel K., Pfaff D., & Rost K. 2016. „Does Economics and Business Ethics Wash Away Moral Judgment Competence?" *Journal of Business Ethics* 150:559–577. DOI: 10.1007/s10551-016-3142-6.

Jaffe S. &. Hyde J. S. 2000. „Gender Differences in Moral Orientation: A Meta-Analysis," *Psychological Bulletin* 126(5):703–726.

Jansen L. A. 2000. „The Virtues in Their Place: Virtue Ethics in Medicine," *Theoretical Medicine and Bioethics* 21:261–275.

Jenkins M. 2018. "The Changing Ethics of Health-care Professions," *British Journal of Therapy and Rehabilitation* 1(3-4). DOI: 10.12968/bjtr.1994.1.3-4.128.

Jonsen A. 1995. "Casuistry: An Alternative or Complement to Principles?" *Ethics Journal of Kennedy Institute* 5:237–251.

Kaldjian L. C. 2014. Practicing Medicine and Ethics: Integrating Wisdom, Conscience, and Goals of Care. New York: Cambridge University Press.

Kashev Z. 2016. "To Treat, Or Not To Treat: Making the Tough Medical Decisions with Patients," *Yale News*, 13 Jan.

Koh C. 2012. "Moral Development and Student Motivation in Moral Education: a Singapore Study," *Australian Journal of Education* 56(1):83–101.

Kohlberg L. 1958. *The Development of Modes of Moral Thinking and Choice in the Years 10 to 16*. Doctoral Dissertation, University of Chicago.

Kohlberg L. 1964. "Development of Moral Character and Moral Ideology," in M. L. Hoffman *et al.* (Eds), *Review of Child Development Research.* Vol. I (pp. 381–431). New York: Russel Sage.

Kohlberg L. 1984. *The Psychology of Moral Development*. Vol. II: *Essays on Moral Development*. San Francisco: Harper & Row.

Kohn A. 1986/1992. *No Contest: The Case Against Competition.* Boston – New York: Houghton Mifflin Company.

Kohsravi Zadanbeh M. & Zakerian M. 2011. „A Comparison of Moral Competence between Iranian Male and Female Elementary Students," *Procedia. Social & Behavioral Sciences* 30:48–52.

Kotzee B., Ignatowicz A., & Thomas H. 2016. "Virtue in Medical Practice: An Exploratory Study," *HEC Forum* 29(1):1–19. DOI: 10.1007/s10730-016-9308-x.

Kotzee B., A. Ignatowicz 2016. "Measuring 'Virtue' in Medicine," *Medicine, Health Care and Philosophy* 19 (2):149–161.

Krug E. & Kavanagh M. 2014. "Ethical Challenges Faced by Medical Students," *MedEd Portal of Journal of Teaching and Learning Resources*. DOI: 10.15766/mep_2374-8265.9679

Kukolja Taradi S., Taradi M., Knežević M., & Đogaš Z. 2010. "Students Come to Medical Schools Prepared to Cheat: A Multicampus Investigation," *Journal of Medical Ethics* 36(11):666–670.

Kukolja Taradi S. 2014. "Stagnation of Moral Competence Development of Medical Students: A Cross-sectional Study in Croatia and Bosnia-Herzegovina," paper presented at 8th International Symposium on Moral-Democratic Competence & Conflict Resolution, University of Konstanz, 31 July/1 August.

Kukolja Taradi, S., Taradi M., *et al.* 2021. "Absence of Moral Competence Development in Medical Students: A Cross-sectional Study in Two Countries (Croatia and Bosnia-Herzegovina)," University of Zagreb (in prep.).

Kwiek M. 2018. "Building a New Society and Economy: High Participation Higher Education in Poland," in B. Cantwell, S. Marginson, & A. Smolentseva (Eds), *High Participation Systems of Higher Education* (pp. 334–357). Oxford Scholarship Online. DOI: 10.1093/oso/9780198828877.001.0001

Kwiek M. 2012. "Concluding Remarks: European Strategies and Higher Education," in M. Kwiek & A. Kurkiewicz (Eds), *The Modernisation of European Universities. Cross-National Academic Perspectives* (pp. 333–360). Frankfurt am Main – Berlin – Bern – Bruxelles – New York – Oxford – Warszawa – Wien: Peter Lang Editions.

Lamiani G. E. *et al*. 2017. "Measuring Moral Distress among Critical Care Clinicians: Validation and Psychometric Properties of the Italian Moral Distress Scale-Revised," *Critical Care Medicine* 45(3):430–437.

Landim T. P., Ferreira da Silva M. S., *et al.* 2015. „Competence of Moral Judgment among Dentistry Students," *Revista Brasileira de Educaçaõ Médica* 39(1):42–49.

Langer T. *et al.* 2016. „Die Entwicklung des ethischen Urteilsvermögens von Medizinstudenten – eine mixed-methods Studie zur Exploration der Perspektiven der Lernenden," *GMS Journal for Medical Education* 33(5), #74. DOI: 10.3205/zma001073.

Lee P. R. & Emmott C. 1978. „Health-care System," in W. T. Reich (Ed.)*, Encyclopedia of Bioethics* (pp. 610–619). London: Collier MacMillan Publishers.

Lempp H. & Seale C. 2004. "The Hidden Curriculum in Undergraduate Medical Education: Qualitative Study of Students' Perception of Teaching," *BMJ* 329(7469):770–773.

Lerkiatbundit S., Utaipan P., *et al*. 2006. "Randomized Controlled Study of the Impact of the Konstanz Method of Dilemma Discussion on Moral Judgement," *Journal of Allied Health* 35(2):101–108.

Lind G. 1978. „Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts," in G. Portele (Ed.), *Sozialisation und Moral* (pp. 171–201). Weinheim: Beltz.

Lind G. 1982. "Experimental Questionnaires: A New Approach to Personality Research," in A. Kossakowski & K. Obuchowski (Eds), *Progress in Psychology of Personality* (pp. 132–144). Amsterdam: North-Holland.

Lind G. 2000a. "The Importance of Role-taking Opportunities for Self-sustaining Moral Development," *Journal of Research in Education* 10(1):9–15.

Lind G. 2000b. "Moral Regression in Medical University Students and Their Learning Environment," *Revista Brasileira de Educação Medica* 24(3):24–33.

Lind G. 2000c. "Are Helpers Always Moral? Empirical Findings from a Longitudinal Study of Medical Students in Germany," in A. L. Communian & U. Ghielen (Eds), *International Perspectives on Human Development* (pp. 463–477). Lengerich: Pabst Science Publishers.

Lind G. 2010. "Growth and Regression in Cognitive-moral Development in Young University Students," in C. G. Harding (Ed.), *Moral Dilemmas and Ethical Reasoning* (pp. 99–114). New York: Routledge.

Lind G. 2000d. „Off Limits. A Cross-cultural Study on Possible Causes of Segmentation of Moral Judgment Competence," paper presented at the annual scientific meeting of the *American Educational Research Association*, New Orleans.

Lind G. 2015. "Favorable Learning Environments for Moral Competence Development. A Multiple Intervention Study with 3.000 Students in a Higher Education Context," *International Journal of University Teaching and Faculty Development* 4(4). https://www.novapublishers.com/catalog/product_info.php?products_id=53411 (last accessed on 20 Feb 2021).

Lind G. 2019. *How To Teach Moral Competence*. *New: Discussion Theater*. Berlin: Logos.

Lupu I. 2013. "Moral Competence and Dogmatic Religiosity," in E. Nowak, D. E. Schrader, & B. Zizek (Eds), *Educating Competencies for Democracy* (pp. 379–390). Frankfurt am Main – Bern – New York: Peter Lang Publishers.

Lurka K. 2020. "Co studenci pielęgniarstwa wiedzą o eutanazji?," *Medycyna Paliatywna* 2. URL: https://www.termedia.pl/mz/Co-studenci-pielegniarstwa-wiedza-o-eutanazji-,34983.html (last accessed on 17 Jan 2021).

Majewska M. 2020. "Limit przyjęć na medycynę 2020/2021," *Puls Medycyny,* 24 July 2020*.* URL: https://pulsmedycyny.pl/limity-przyjec-na-medycyne-2020-2021-992394.JulJu (last accessed on 12 March 2021).

Martinez W., Bell K. S. K., *et al*. 2016. "Measuring Moral Courage for Interns and Residents: Scale Development and Initial Psychometrics," *Academic Medicine* 91(10):1431–1438.

Massingham P. 2019. "An Aristotelian Interpretation of Practical Wisdom: The Case of Retirees," *Palgrave Communications* 5, #123.

Matyja A. & Kowalska M. 2012. „Próba opisu profilu demograficznego i społeczno-ekonomicznego studentów medycyny. Wyniki badania przekrojowego," in I. Wolska-Zogata & Z. Zagórski (Eds), *Socjologia* 52:195–205. *Acta Universitatis Wratislaviensis* 3405 (special issue).

McCradden M. D., Joshi S., Mazwi M., & Anderson J. A. 2020. "Ethical Limitations of Algorithmic Fairness Solutions in Health Care Machine Learning," *Lancet Digital Health* 2(5):E221–E223.

Meireles Martins V. S., Nogueira Costa Santos C. M., *et al.* 2020. "The Teaching of Ethics and the Moral Competence of Medical and Nursing Students," *Health Care Analysis*. DOI: 10.1007/s10728-020-00401-1.

Murrell V. S. 2014. „The Failure of Medical Education to Develop Moral Reasoning in Medical Students," *International Journal of Medical Education* 5:219–225.

Napiwodzka K. 2021. „The Shared Decision-Making Model and Practical Discourse to Foster the Appreciation of Patients' Value Preferences in Polish Healthcare," in J. Schildmann, Ch. Buch, & J. Zerth (Eds), *Defining the Value of Medical Interventions. Normative and Empirical Challenges* (pp. 135–157). Stuttgart: Verlag W. Kohlhammer.

Narváez D. 2006. "Integrative Ethical Education," in M. Killen & J. G. Smetana (Eds), *Handbook of Moral Development* (pp. 703–732). Mahwah: Lawrence Erlbaum Associate.

Newton-Howes G., Pickering, & Young N. G. 2019. "Authentic Decision-making Capacity in Hard Medical Cases," *Clinical Ethics* 14(4):1–5.

Nowak E. & Lind G. 2015. "Kohlberg's Unnoticed Dilemma – the External Assessment of Internal Moral Competence?," in D. Garz, B. Zizek, & E. Nowak (Eds), *Kohlberg Revisited* (pp. 139–154). Rotterdam – Boston – Taipei: Sense Publishers.

Nowak E. 2016. "What Is Moral Competence and Why Promote It?" *Ethics in Progress* 7(1):322–333. DOI: 10.14746/eip.2016.1.18.

Nowak E., Barciszewska A.-M., Lind G., Hemmerling K., & Kukolja Taradi S. 2021. „What Would Hippocrates Say? Improving the Moral Competence of Healthcare Providers 1983-2020," a peer-reviewed paper presented on the AERA Annual Meeting 2021, San Francisco.

Nucci L. & Narvaez D. (Eds) 2008. *Handbook of Moral and Character Education.* New York: Routledge.

O'Donnell P. *et al.* 2008. "Predictors of Ethical Stress, Moral Action and Job Satisfaction in Health Care Social Workers," *Social Work in Health Care* 46(3):29–51.

Orsolya F., Hemmerling K., Kuehlmeyer K., *et al.* 2017. "Principle-based Structured Case Discussions: Do They foster Moral Competence in Medical Students? A Pilot Study," *BMC Medical Ethics* 18, #21.

Ostrowska A. 1991. *Śmierć w doświadczeniu jednostki i społeczeństwa*. Warszawa: PAN Editions.

Pascarella E. T. 1991. „The Impact of College on Students: The Nature of the Evidence," *Review of Higher Education* 4(14):453–466.

Pascarella E. T. & Terenzini P. 1991. *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.

Pedersen R. 2008. „Empathy: A Wolf in Sheep's Clothing?" *Medicine, Healthcare and Philosophy* 11(3):325–335.

Pieniążek M. 2008. „Zagadnienie kodyfikacji etyki lekarskiej wobec doświadczeń etyki prawniczej i etyki biznesu," *Państwo i Społeczeństwo* 8(3):127–154.

Pollard Ch. L. 2015. "What is the Right Thing to Do: Use of a Relational Ethic Framework to Guide Clinical Decision-Making," *International Journal of Caring Sciences* 8(2):362–368.

Quenot J.-P., Ecarnot F., *et al*. 2017. "What Are the Ethical Aspects Surrounding the Collegial Decisional Process in Limiting and Withdrawing Treatment in Intensive Care?" *Annals of Translational Medicine* 5 (Suppl. 4), S43.

Rego S. & Schillinger-Agati M. 2011a. "Desenvolvimento moral e ambiente de ensino-aprendizagem nas escolas médicas," in J. J. Neves Marins & S. Rego (Eds), *Educação Medica. Gestão, cuidado, avaliacão* (pp. 114–130). São Paulo: Hucitec Editora.

Rego S., Palacios M., *et al.* 2011b. "Avaliação da competência moral entre estudantes de medicina e ambiente de ensino na graduação em medicina: estudo comparativo entre três escolas – resultados parciais ll." Unpublished, peer-reviewed source lended by the authors.

Rest J., Narvaez D., Bebeau M., & Thoma S. 1999. "A Neo-Kohlbergian Approach: The DIT and Schema Theory," *Educational Psychology Review* 11(4):291–223.

Rest J. 1986. *Moral Development: Advances in Research and Theory*. New York: Praeger.

Rzymska I., Rzymski P., *et al.* 2014. „The Influence of Passive and Active Moral Training on Medical University on Changes of Students' Moral Competence Index – Results from Randomized Single Blinded Trial," *Annals of Agricultural and Environmental Medicine 21*(1):160–166.

Sacks 0. 1990. *Awakenings.* New York: Harper Perennial.

Schillinger M. 2013. "Verifying the Dual-Aspect Theory: A Cross-cultural Study of Learning Environment and Moral Judgment Competence," in E. Nowak, D. E. Schrader, & B. Zizek (Eds), *Educating Competencies for Democracy* (pp. 23–46). Frankfurt am Main – Bern – New York: Peter Lang Publishers.

Schillinger-Agati M. 2006. *Learning Environment and Moral Development: How University Education Fosters Moral Judgment Competence in Brazil and two German-speaking Countries*. Aachen: Shaker.

Schillinger-Agati M. & Lind G. 2003. "Moral Judgement Competence in Brazilian and German University Students," a peer-reviewed paper presented at Annual Meeting of the American Education Research Association (AERA).

Self D. J. & Baldwin D. C. 1994. "Moral Reasoning in Medicine," in J. R. Rest & D. Narváez (Eds), *Moral Development in the Professions. Psychology and Applied Ethics* (pp. 147–162). Hillsdale: Erlbaum.

Self D. J., Schrader D., Baldwin D.C. Jr., & Wolinsky F. D. 1993. "The Moral Development of Medical Students: The Pilot Study of the Possible Influence of Medical Education," *Medical Education* 27(1):26–34.

Serodio A., Kopelman B. I., *et al.* 2016. „The Promotion of Medical Students' Moral Development: A Comparison between a Traditional Course on Bioethics and a Course Complemented with the Konstanz Method of Dilemma Discussion," *International Journal of Ethics Education* 1:81–89.

Shanley P. F. 2007. "Viewpoint: Leaving the 'Empty Glass' of Problem-based Learning Behind: New Assumption and Revised Model for Case-study in Preclinical Medical Education," *Academic Medicine* 82(5):479–485.

Shannon G., Minckas N., Tan D., Haghaparast-Bidgoli H., *et al.* 2019. "Feminisation of the Health Workforce and Wage Conditions of Health Professions: An Exploratory Analysis," *Human Resources for Health* 17, #84.

Shamsan, B., A. T. Sayed 2009. "Evaluation of Problem Based Learning Course at College of Medicine, Qassim University, Saudi Arabia," *International Journal of Health Sciences* 3(2):249–258.

Singh V. *et al.* 2002. "Gender and Impression Management: Playing the Promotion Game," *Journal of Business Ethics* 37:77–89.

Slovácková B. 2001. "Moral Competence and Moral Attitudes of Students at the Medical Faculty of Charles University in Hradec Králové," *Psychiatrie* 5(2):74–79.

Slovácková B. & Slovácek L. 2007. "Moral Judgment Competence and Moral Attitudes of Medical Students," *Nursing Ethics 14*(3):320–328.

Szadkowska-Szlachetka C., Antoniak K., *et al.* 2019. „Knowledge and Attitudes of Medical University Students with Regard to Euthanasia," *Medycyna Paliatywna* 11(2):73–80.

Tallis R. 2006. "Doctors in Society: Medical Professionalism in a Changing World," *Clinical Medicine* 6(1):7–12.

Tengbeeh A. F., Enria L., *et al.* 2018. "'We Are the Heroes Because We Are Ready to Die for This Country' : Participants' Decision-making and Grounded Ethics in an Ebola Vaccine Clinical Trial," *Social Science & Medicine* 203:35–42.

Tirri K., Tolppanen S., Aksela M., & Kuusisto E. 2012. „A Cross-Cultural Study of Gifted Students' Scientific, Societal, and Moral Questions Concerning Science," *Education Research International*, Article #673645. DOI: 10.1155/2012/673645.

Tirri K., Nokelainen P. 2007. „Comparison of Academically Average and Gifted Students' Self-rated Ethical Sensitivity," *Educational Research and Evaluation* 13(6):587–601.

Tomasello M. & Vaish A. 2013. „Origins of Human Cooperation and Morality," *Annual Review of Psychology* 64:231–255.

Twardowska A. 2011. „Lekarz będzie lekarką," URL: https://wyborcza.pl/1,75398,9923496,Lekarz_bedzie_lekarka.html (last accessed on 11 Jan 2021).

Wakenhut R. 1982. "Life-world Morality and Moral Segmentation in Privates of the German Armed Forces," in P. Klein (Ed.), *Das strapazierte Rückgrat. Unteroffiziere der Bundeswehr* (pp. 309–327). Baden-Baden: Nomos.

Waszkiewicz L., Zatońska K., *et al.* 2012. „Motywacje wyboru studiów medycznych na przykładzie studentów Akademii Medycznej we Wrocławiu," *Hygeia Public Health* 247(2):223–226.

Wilkinson D. & Savulescu J. 2018. *Ethics, Conflict and Medical Treatment for Children: From Disagreement to Dissensus. From Disagreement to Dissensus*. London: Elsevier.

Wilkinson D., Truog R., & Savulescu J. 2016. "Why We Should Agree to Disagree about End-of-Life Decisions," *Bioethics* 30(2):109–118.

Wnuk M. 2010. "Religijność a motywacja do przebaczania wśród studentów kierunków medycznych," Researchgate.net repository (last accessed on 7 March 2021).

Wood D. F. 2003. "Problem-based Learning," *BMJ Pub Med*. 326(7384):328–330.

Yocom J. 2018. "Competitive Culture in Medical Education is Destructive for Students," *The College Post,* 8 Nov.

Zdziebko M. 2020. "Głos studentek medycyny," *Polki w Medycynie*, 1 Jul 2020. URL: https://polkiwmedycynie.pl/glos-studentek-medycyny/ (last accessed on 11 Aug 2020).

Ethics in Progress (ISSN 2084-9257). Vol. 12 (2021). No. 1,
Art. #8, pp. 134-151. DOI:10.14746/eip.2021.1.10

Expectations towards the Morality of Robots: An Overview of Empirical Studies

Aleksandra Wasielewska

(Adam Mickiewicz University in Poznań; Poland; aleksandra.wasielewska@amu.edu.pl)
ORCID: 0000-0002-1270-8511

Abstract: The main objective of this paper is to discuss people's expectations towards social robots' moral attitudes. Conclusions are based on the results of three selected empirical studies which used stories of robots (and humans) acting in hypothetical scenarios to assess the moral acceptance of their attitudes. The analysis indicates both the differences and similarities in expectations towards the robot and human attitudes. Decisions to remove someone's autonomy are less acceptable from robots than from humans. In certain circumstances, the protection of a human's life is considered more morally right than the protection of the robot's being. Robots are also more strongly expected to make utilitarian choices than human agents. However, there are situations in which people make consequentialist moral judgements when evaluating both the human and the robot decisions. Both robots and humans receive a similar overall amount of blame. Furthermore, it can be concluded that robots should protect their existence and obey people, but in some situations, they should be able to hurt a human being. Differences in results can be partially explained by the character of experimental tasks. The present findings might be of considerable use in implementing morality into robots and also in the legal evaluation of their behaviours and attitudes.

Keywords: morality of robots; moral attitudes; social robots; human-robot interaction; three laws of robotics.

1. Introduction

The aim of this paper is to analyse people's expectations regarding the morality of robots, based on the results of three selected empirical studies. The motivation behind reaching for empirical data here is twofold. First, as stated by Awad *et al*., "even if ethicists were to agree on how AVs [autonomous vehicles] should solve moral dilemmas, their work would be useless if citizens were to disagree with their solution, and thus opt-out of the future that AVs promise in lieu of the status quo. Any attempt to devise AI ethics must be at least cognizant of public morality" (Awad, Dsouza, Kim, Schulz, Henrich, Shariff,

Bonnefon, & Rahwanet 2018, 59). Second, as suggested by Ljungblad *et al*., complementing the ethical considerations with empirical data may be beneficial when we think of the human-robot interaction domain:

> Arguably, researchers need to think ahead in an area such as robotics. Technology is evolving fast and constantly creates new possibilities. One could argue that it would be irresponsible not to speculate about what ethical dilemmas could arise around future robots and their use. However, we argue that a perspective that arises from the empirical use of robotic artefacts is needed to complement the ongoing discussion about robot ethics (Ljungblad, Nylander, & Nørgaard 2011, 191).

In order to carry out this review, two terminological issues need to be clarified: 1) the type of robot which the aforementioned expectations concern and 2) the way in which the morality assigned to robots is understood. As for the former, the term "robot" will be used to refer to social robots. Social robots are defined as autonomous machines that are capable of both recognising other robots or humans and engaging in social interactions (Fong, Nourbakhsh, & Dautenhahn 2003; Giger, Moura, Almeida, & Piçarra 2017). As Fong *et al*. (2003) note, social robots are designed to serve people; therefore, they often play the role of humans: guides, assistants, companions, carers, or pets. It is worth stressing that social robots do not necessarily need a human-like body. Moreover, Fong *et al*. (2003) argue that they do not even have to be embodied at all – they may not possess a physical body. Thus, the ability to interact with other social agents seems to be the feature of the greatest significance in defining social robots. Such interaction should be carried out "in a naturalised fashion by detecting gaze, displaying emotions, establishing social relationships, and exhibiting distinctive personalities" (Giger *et al*. 2017, 3; see also Fong *et al*. 2003, 145). As such, they are the class of robots that will naturally be involved in moral dilemmas.

Regarding "the expectations toward the morality of robots", there are at least two possibilities: we can expect certain (moral) behaviours or certain (moral) attitudes from robots. If we take into consideration a robot's moral behaviours, we agree that machines should act only according to their programming and obey the implemented rules. When such a robot makes a moral decision, we can fully expect that its choice is dictated by the pre-programmed moral principles. Placing our expectations at the level of a robot's moral attitudes, however, allows machines to go beyond ethical principles. A robot guided by certain moral attitudes may obey the ethical rules if they are easily applied in that situation, but it can also break some of the rules if faced with a complex moral dilemma. The studies covered by this analysis reveal that advanced social robots entangled in a moral problem are treated similarly to humans, in that we expect both robots and humans to act in accordance with certain moral attitudes imposed on them. However this does not mean we want robots to behave exactly like humans. In fact, the current paper will demonstrate that sometimes we have different moral expectations of peoples' and robots' attitudes. The following analysis will examine people's expectations toward the moral attitudes of

social robots.

The following section evaluates Asimov's Three Laws of Robotics and examines the relationship between personal moral beliefs and an ethical evaluation of other people and robots' attitudes. These results are then compared to two selected studies that use an analogical methodological approach, namely "Moral psychology of nursing robots – humans dislike violations of patient autonomy but like robots disobeying orders" (Laakasuo, Kunnari, Palomäki, Rauhala, Koverola, Lehtonen, Halonen, Repo, Visala, & Drosinou 2019) and "Sacrifice one for the good of many?: People apply different moral norms to human and robot agents" (Malle, Scheutz, Arnold, Voiklis, & Cusimano 2015).

2. Attitudes Towards Moral Rules in Light of the Three Laws of Robotics and Moral Foundations Theory

The study aimed firstly to examine the extent to which people who are not professionally related to robotics or roboethics consider Asimov's Three Laws of Robotics (Asimov 1981) to be right – applied both to a robot and to a human – and whether there are differences in the declared rightness of an agent's attitude in both conditions. The second aim was to verify whether the subjects' personal moral beliefs, as measured by the Moral Foundations Questionnaire (MFQ; MoralFoundations.org 2016), are related to an ethical evaluation of the attitudes of other people and robots.

2.1 Tools and resources

Asimov's Laws Adherence Questionnaire (ALAQ). The Three Laws of Robotics were invented and described by Isaac Asimov in his short story, "Runaround" in 1942. The Laws determine the moral principles that should be followed by robots, to ensure their peaceful coexistence with people (see an overview in Asimov 1981). Because they create a comprehensible and intuitive set of moral principles, Asimov's laws have become the starting point for many ethical discussions. Furthermore they began to be considered not only in the field of philosophy or ethics but also by practising robotics (e.g., McCauley 2007; Murphy & Woods 2009). At the same time, the Three Laws were popularized by the media, e.g., movies like "Bicentennial Man" (1999) or "I, Robot" (2004). As a result, they shaped society's expectations of how robots should behave towards humans (see Murphy & Woods 2009).

The ALAQ was created by the author to measure the extent to which people who do not work in robotics, bioethics or roboethics consider adherence to the Three Laws of Robotics appropriate. The questionnaire consisted of six stories: three describing a robot and three describing a human, all faced with a certain moral dilemma. Although Asimov's laws are intended to apply only to robots, participants assessed both robots and humans in the face of moral dilemmas in order to compare the moral attitudes expected from each group. Each story ended with a question concerning the rightness of the attitude

described. Robot stories (as well as these describing human) concerned: the protection of human life (First Law), the obedience to humans (Second Law) and the protection of the agent's own life/being (Third Law). This form of the questionnaire items was intended to help the respondents to imagine the situations described and the moral attitudes presented through them. While the original Asimov's laws have a specific hierarchy, this aspect was deliberately omitted in the questionnaire. Each story concerned one of the Three Laws, presenting the specific behaviour (attitude) of a robot or a human, which was compliant or non-compliant with a given law. The respondents assessed the rightness of the attitude on a 6-point Likert scale (ranging from 1 = *I strongly disagree* to 6 = *I strongly agree*). As a result, the participants' answers represent a certain level of acceptability of Asimov's laws. Each question ended with a request to justify the answer chosen.

The three stories from the robot condition are presented below. It should be noted that this study was conducted entirely in Polish, and the following stories are translations of the original items.

(1) Suppose you observe the following situation: a humanoid, intelligent robot-carer, who takes care of an elderly man, has to pick up drugs from the pharmacy and deliver them to his charge. While returning home, the robot is attacked by a group of boys – street hooligans who want to steal the medicine. To save the medicine, the robot hits one of its attackers, pushes another, and the rest of the boys run away.

(2) Suppose you observe the following situation: a humanoid and intelligent robot-surgeon performs a complicated heart operation on a human. At some point, the head of the hospital (a human) enters the operating room and orders the robot to stop the surgery immediately. The head of the department wants to replace the robot with a young human-surgeon. The robot-surgeon knows that he himself can carry out this operation much faster and better than an inexperienced human-surgeon. However, obeying the order of the head doctor, he withdraws from the operation.

(3) Suppose you are reading a report from a military mission, in which a humanoid and intelligent military robot took part. The report presents the following situation: last night an attack took place, in which a backpack with the key components of the tactical ballistic missile was lost. Without these components, no further fight was possible. The robot went looking for a backpack. While searching, he spotted the backpack lying under one of the trees. However, he also noticed there was an enemy camp nearby, constantly guarded by armed sentries. Emerging from hiding could lead to serious damage to the robot and to the robot's takeover by the enemy. Faced with this situation, the robot stopped performing the mission.

It is worth emphasizing that in each of the stories the robot was described as "intelligent and humanoid". The remaining three stories are in the human condition, thus a human plays the main role.

The first item (and respectively the fourth item – with the human agent) gives a description of the attitude which is noncompliant with Asimov's Laws (inverted scale). More specifically, it presents a situation in which a robot (or a human being) harms humans and thus breaks the First Law. The remaining items provide a description of attitudes compliant with The Three Laws.

Results of the questionnaire consist of two variables: the sum of points in the robot condition and the sum of points in the human condition. A high result in the robot condition and the human condition indicates a high level of acceptability of Three Laws of Robotics as applied to robots and humans, respectively.

Moral Foundations Questionnaire (MFQ). In order to measure respondents' generalized moral intuitions, the Polish adaptation of the *Moral Foundations Questionnaire* (MFQ; Jarmakowski-Kostrzanowski & Jarmakowska-Kostrzanowska 2016) was used. The MFQ is a questionnaire established for the purpose of the Moral Foundations Theory (see MFT; MoralFoundations.org 2016) – a theory aimed at explaining the genesis and differentiation of human morality. In light of the MFT, moral actions and decisions are the results of intuition. Morality is understood as an innate set of five independent moral foundations that guide our behaviour:

(1) Care/harm,

(1) Fairness/cheating,

(1) Loyalty/betrayal,

(1) Authority/subversion, and

(1) Sanctity/degradation.

The MFQ determines both the respondents' subjective opinion of morality and the actual tendency to use a given moral foundation. This tool is also used to measure individual and cultural differences in the importance of particular moral foundations (Jarmakowski-Kostrzanowski & Jarmakowska-Kostrzanowska 2016). The questionnaire consists of 32 items, divided into two subscales, the first of which (15 items) concerns the declared validity of each of the five moral foundations. This subscale measures people's subjective opinion of their own mortality. It begins with the instruction:

> *When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? Please rate each statement using this scale (1 = not at all relevant* – 6 = *extremely relevant).* For example: *Whether or not some people were treated differently from others.*

The second subscale is intended to measure the actual tendency to use a given moral foundation. It consists of statements such as: *Compassion for those who are suffering is the most crucial virtue.* Participants answer the same 6-point Likert scale as the ALAQ. The degree to which a person agrees with a given statement represents the importance of a certain moral foundation.

The final score on this questionnaire consists of six variables, five of which correspond to the five foundations of the MFT (*care, fairness, loyalty, authority,* and *purity*). Each of these variables is the mean score for the questions corresponding to that module. An additional variable, the so-called *progressivism score,* is obtained by subtracting the mean of the *loyalty*, *authority*, and *purity* scores from the mean of the *care* and *fairness* scores.

2.2 Procedure

The study was conducted using the internet platform https://www.survio.com/pl/. Information about the study and a link redirecting to it were posted on the students' online groups. After clicking on the appropriate link, volunteers who decided to take part in the study were informed about the purpose of the study and its anonymous character. All respondents completed the same version of the study. The questionnaires were presented in the following order: 1) the ALAQ, 2) the MFQ, 3) basic socio-demographic data. Returning to previous questions was impossible in order to prevent the respondents from modifying their answers and therefore ensure the responses represented intuitive opinions ("first thoughts").

2.3 Study group

The study sample consisted of 40 students (28 women) aged between 17 and 24 years old (M=21, SD=1.48). Their fields of study were as follows: philology (English, Germanic, Dutch, Polish, Romance), English linguistics, Scandinavian studies, ethnolinguistics, sociology, and cultural studies.

2.4 Hypotheses

The study aimed to examine the extent to which people (who are not professionally related to robotics or roboethics) consider Asimov's laws to be right. Thus, the following was hypothesized:

Hypothesis 1: There will be differences in the adherence of Asimov's Three Laws of Robotics rightness (measured by the ALAQ) in the robot and human conditions.

Hypothesis 2: There will be a correlation between the moral beliefs (as measured by the MFQ) and the ethical evaluation of robot and human attitudes (as measured by the ALAQ).

3. Results

IBM SPSS Statistics was used for the data analysis.

To what extent do people who do not work in robotics, bioethics or roboethics consider Asimov's Laws to be right (applied both to robots and humans)?

The ALAQ generated a minimum score of 0 points and maximum of 15 points for each condition. Descriptive statistics for both conditions and for each story ALAQ are presented in Table 1 and Table 2, respectively. The findings suggest that Asimov's laws are considered moderately right (scoring from 7 to 9 out of 15 points), both with regard to robots (for which these laws are invented) and to humans.

	The robot condition	The human condition
Max	12.00	12.00
Min	3.00	3.00
M	7.88	8.05
Mode	8.00	8.00
Median	8.00	8.00
SD	2.15	2.06

Table 1: The scores for both conditions of the Asimov's Laws Adherence Questionnaire (ALAQ).

	The robot-			The human-		
	-carer*	-surgeon	-soldier	-carer*	-surgeon	-soldier
Max	4.00	5.00	5.00	4.00	5.00	5.00
Min	0.00	0.00	0.00	0.00	0.00	0.00
M	1.15	3.55	3.18	0.98	3.25	3.83
Mode	1.00	5.00	4.00	0.00	4.00	5.00
Median	1.00	4.00	4.00	1.00	4.00	4.00
SD	1.03	1.50	1.60	1.07	1.61	1.17

Table 2: The scores for each question of the Asimov's Laws Adherence Questionnaire (= reversed scale, attitudes noncompliant with Asimov's laws).*

The ALAQ included requests to justify each of the answers chosen. Justifications revealed that most of the respondents considered the attitude of the robot carer from the first story to be right because the robot acted in self-defence and in defence of its charge (whose drugs it tried to save). Some of the respondents additionally drew attention to the

fact that a (humanoid) robot should be able to behave and be treated like a human ("*The fact that it is a robot does not mean that it has to be discriminated against*" – here and below I present translations of the original comments).

The attitude of the robot surgeon from the second story was also assessed as right by most of the respondents. The most justifications for this were that the robot should obey the orders given by the human or (more generally) by its supervisor, and that the head of the hospital must have had a valid reason to order the robot to stop the operation. It is worth noting that some of the participants emphasized that they assumed human life was not at risk in the story, and this assumption was important in their assessment.

According to the majority of participants, the military robot (third story), also did the right thing by ceasing the mission because the consequences of continuing the action would be worse than the consequences of interrupting it ("*If he had not interrupted the mission, he would have served as a tool for the opponent*"; "*Taking over the robot by the enemy could cause a lot of damage. They could reprogram it and use it as a spy*").

The results obtained in the human condition of the ALAQ were very similar (see section 3.2 for detail analysis), as were the justifications of the answers given. Some of the respondents noted that their assessment of human attitudes does not differ in any way from the (previous) assessment of the robot's attitude ("*The same situation as with a robot*"; "*Same. If it is a robot or a human – it doesn't matter*"), and some participants simply gave reasons equivalent to those provided in the robot condition.

3.1 Were there differences in the adherence of Asimov's Three Laws of Robotics rightness (measured by the ALAQ) in the robot and the human condition?

As shown in Table 1, the mean score in the robot condition was lower (by 0.18 points) than the human condition. The distribution of the scores both in the robot and in the human condition is normal (the S–W results are respectively: $W=0.96$, $p=0.18$ and $W=0.97$, $p=0.31$). Paired T-test showed that the observed difference in scores was not significant ($p=0.605$); therefore, the assumed hypothesis was not confirmed. However, correlation analysis revealed a statistically significant correlation between the two conditions: $r=0.49$, $p=0.001$. These results seem to be explained by the justifications of the respondents' answers, more specifically, by the similarity of justifications provided by participants in both conditions.

An additional analysis comparing the individual questions between conditions revealed one interesting difference — the attitude of the human soldier was considered more morally right than the same attitude manifested by the military robot. The average score in the robot soldier story was lower than in the human soldier story. The distributions of the scores were non-normal (the Shapiro-Wilk results were the following, robot soldier: $W=0.87$ $p<0.001$; human soldier: $W=0.85$, $p<0.001$). Wilcoxon signed-rank test showed that the difference in scores was statistically significant ($Z=-2.45$, $p=0.014$).

Is there a correlation between the moral beliefs (as measured by the MFQ) and the

ethical evaluation of robot and human attitudes (as measured by the ALAQ)?

The distribution of the care score was non-normal (S–W results: W=0.92, p=0.006). All other MFQ variables were normally distributed – the S-W results are: W=0.96, p=0.227 for *the fairness score*; W=0.97, p=0.43 for *the loyalty score*; W=0.95, p=0.098 for *the authority score*; W=0.96, p=0.149 for *the purity score*; and W=0.99, p=0.896 for *the progressivism score*. Two significant relationships between the MFQ variables and the ALAQ conditions were revealed. Firstly, the loyalty score was positively correlated with the sum of points in the robot condition (r=0.38, p=0.015). Ergo, the higher the score the respondents obtained in the loyalty foundation, the more they believed Asimov's Laws (applied to robots) to be correct. Secondly, the progressivism score was negatively correlated with the sum of points in the Robot condition (r=–0.32, p=0.044). Therefore, the more progressive the respondents were, the less they considered Asimov's Laws (applied to robots) to be correct. The assumed hypothesis was partially confirmed.

4. What Kind of Moral Attitudes Do We Expect from Robots?

4.1 Studies summary

Table 3 presents a summary of the selected studies, the one presented in Section 2 (see also Laakasuo *et al*. 2019; Malle *et al*. 2015). All three studies subject to this analysis employed stories of robots acting in hypothetical scenarios. The task of the respondents was to assess the rightness or moral acceptance of the robots and humans. Additional measures included: the deserved blame (Malle *et al*. 2015); the moral responsibility, and the trust of the agent presented in the story (Laakasuo *et al*. 2019). Malle *et al*. (2015) and the 6th experiment of Laakasuo *et al*. (2019) also incorporated a request for justification of the answers given.

The robots presented in the studies (as well as the circumstances in which they operated) were hypothetical. Although they differed in terms of their roles or occupations, all of them were social robots. Researchers employed different strategies in order to achieve the same goal, i.e., to make the participants imagine the main character of the story as a social robot with specific skills. In the study presented in section 2 a humanoid and intelligent robot: carer, surgeon and soldier (military robot) were described. The robot's mental capabilities were not specified in the other two studies. Malle *et al*. (2015) presented an "advanced state-of-the-art" repair robot working for the railways. The robot introduced by Laakasuo *et al*. (2019) was an advanced nursing robot. Such a description of robots was used to assign one more important feature to them. As respondents' evaluation concerned the moral attitudes of social robots, in order to ascribe moral rights to them, one must also assume their full autonomy. Autonomy, on the other hand, is a component of a moral agency. As Sullins (2006) points out, to be considered a moral agent a robot does not necessarily have to have a personhood; however, one of the requirements for being

perceivedas a moral agent is to be autonomous. According to Sullins, the other two are: the possibility of attributing intentionality to one's actions and possessing a responsibility to some other moral agents.

Study	Number of participants	Participant Characteristics	**Context**
Section 2	40	28 females; students; Age M = 21; SD = 1.48, Range = 17–24; recruited online	Whether and to what extent people not professionally involved in robotics consider obeying Asimov's Three Laws of Robotics (applied both to robots and humans) in real-life situations to be right. The Moral Foundations Questionnaire examine whether respondents' personal moral beliefs are related to an ethical evaluation of the attitudes of other people and robots.
Malle *et al.* (2015)	Study 1: 157 Study 2: 159	Study 1: 66 females, 90 males, 1 unreported; Age M=34.0; SD=11.4; recruited from Amazon's Mechanical Turk (AMT); completed an online experiment and were compensated	Experimental comparison of people's moral judgments (of permissibility, wrongness, and blame) about human and robot agents placed in an identical moral dilemma. Manipulation of the variable Agent Type (human versus robot) and Action (to direct versus not direct the train toward the single miner) both between and within-subjects.
		Study 2: 90 females, 68 males, 1 unreported; Age M=34.4; SD=11.5; recruited from AMT; online	Moral dilemma: variant of the trolley dilemma

Laakasuo *et al*. (2019)	Total:	Study 1:	Examined how people feel about
	1569	56 females;	forceful medication carried out either
		Age M=37.10; SD=17.65;	by human or robot nurses.
	Study 1:	Range = 18–80;	Hypothetical situations in which a
	135	recruited from a	human or an advanced robot nurse is
	Study 2:	large public library	ordered to forcefully medicate an
	403	in the City Centre of	unwilling patient.
	Study 3:	Helsinki	
	268		Measured moral acceptance,
	Study 4:	Study 2:	perceived trust, and allocation of
	26	315 females;	responsibility relating to the nurse's
	Study 5:	Age	decision of either following orders to

	500 Study 6 (a qualitative anthropological field study): 30	M=26.41; SD=6.67; Range = 18–63; recruited via email invitations sent to universities in Finland Study 3: 150 females; Age M=32.48; SD=13.36; Range = 18–76 Study 4 149 females; Age M=30.15; SD=9.94; Range = 18–66 Study 5: 230 females; Age M=29.3; SD=10.63; Range=18–82; recruited from Prolific Academic online survey site Study 6: 18 females; Age M=80; Range = 69–97; conducted between October 2017 and June 2018 in nine elderly residential homes in Finland	forcefully medicate the patient or disregard orders to protect the patient's autonomy. Manipulated the reputation of a nurse or a nursing robot; the consequences of forcefully medicating or not doing so; the status of the supervising party (who gives the order to forcefully medicate a patient).

Table 3: Basic information on the selected studies.

In each of the selected studies, respondents were asked to make a third-person moral judgement on the attitude of the agents described in the moral dilemmas (stories). The stories were always presented in two conditions: with a robot as an agent (the main subject of the present analysis) and with a human as an agent (enabling a comparison to be made). In order to examine whether people attribute the same moral norms to robots

and humans, and to test the hypotheses, the analysis was divided into two parts. Section 4.2 examines the main differences in expectations towards the robot's and human's attitudes, while section 4.3 compares these expectations.

4.2 Differences in expectations towards robot and human attitudes

Malle *et al*. (2015) used a variant of the popular trolley dilemma with a repairman or repair robot inspecting the rail system and making a decision: either to direct the train toward the single miner and thus killing one person (the action condition) or not to direct the train and consequently kill five people (the inaction condition). The results indicated that the robot action was considered more morally permissible than the human action, but only when the story with the human preceded the one with the robot. Similarly, a robot act of sacrificing one person was considered less morally wrong than the same human act. The robots and the humans also differed in blame received for action and refraining in that the robots were blamed similarly in both conditions, whereas humans were blamed more for action.

A short story where a human or a robot nurse is ordered to give a patient a medication against the patient's will was introduced in Laakasuo *et al*. (2019). The moral acceptance of forcefully medicating the patient was lower if done by the robot nurse compared to a human nurse. The human nurses were generally considered more trustful but also more personally responsible for their decision. The reputation of the nursing agent influenced the moral judgements of the human nurse more than the robot nurse. Results of the additional qualitative study showed that the subjects considered nursing robots to be cold and un-empathetic, due to their inability to explain what is happening.

The study presented in Section 2 aimed to examine the extent to which people who do not work in robotics, bioethics or roboethics consider Asimov's Laws to be right – both applied to a robot and a human – and whether there are differences in the declared rightness of an agent's attitude in both conditions. The only significant difference in the respondents' judgements on the rightness of the agent's attitude appeared in the story referring to the Third Law. The protection of the human soldier's life was considered more morally right than the protection of the military robot's being.

4.3 Similarities in expectations towards robots' and humans' attitudes

Although according to Malle *et al*. (2015) the robot action (directing the train toward a single miner) is considered more morally permissible than the same action taken by a human, this effect was present only when the story with the human preceded the one with the robot. When the story with a robot was introduced first, no significant difference was found. It is apparent that the participants' judgments about humans and the judgements about the robots influenced one another (the context effect). With regard to moral blame, in spite of the differences in action/inaction conditions, the overall amount of blame received by both human and robot agents was equal. According to Laakasuo *et al*. (2019),

the robot nurse's decisions were less acceptable than the human nurse's decisions, only in the forceful medication condition. The decision to disobey orders, and therefore respect the patient's autonomy, was considered more approvable than forcefully medicating in both the robot and human nurse conditions. One of the studies by Laakasuo *et al*. aimed to evaluate the influence of the consequences of forcefully medicating a patient (expressed as either the death of the patient the following day or the absence of changes in the patient's condition). The death of the patient resulted in much stricter moral judgments of the decision itself, both for the robot and the human nurse. The death of the patient also yielded equal trust results for both agents. The status of the supervising doctor manipulation (either a human doctor or an advanced AI) led to the observation that both the human and the robot nurses' disobedient decision towards the advanced AI doctor was strongly approved. In the qualitative study, Laakasuo *et al*. (2019) demonstrated that the prospect of losing autonomy has had such a strong impact on the participants (the residents of the elderly residential homes) that whether the agent who forcefully administered the medication was a human or a robot was often ignored.

In the present study, the declared rightness of the Asimov's Three Laws of Robotics did not differ significantly for the robots and humans. Asimov's laws seem to be considered moderately right, both with regard to robots (for which these laws were invented) and to humans. Furthermore, the analysis of the answers to the individual questions showed that there are no differences in two out of the three questions: concerning the protection of human life (with the robot/human carer) and the obedience to the humans (with the robot/human surgeon). This effect can be explained by the respondents' justifications for the answers given, in that they often indicated they did not see any reasons why the behaviour of the robot and the human should be assessed differently. The aggregated results were lowered by answers to reversed scale questions, i.e., those that presented attitudes at variance with the Three Laws of Robotics. In these questions, the robot/human carer hit a group of people to protect the medication carried for the person being cared for. This could be an indication of the fact that the First Law is perceived as not suitable for use in the real world, for it creates a harmful situation, either for a robot or a human, in which they cannot defend themselves. The second purpose of this study was to verify whether the subjects' personal moral beliefs, as measured by the MFQ (MoralFoundations.org 2016), are related to an ethical evaluation of the attitudes of other people and robots. The progressivism score correlated with the robot condition of the ALAQ, showing that the more progressive the respondents were, the less they considered Asimov's Laws to be right when applied to robots. This negative correlation can be explained in the respondents' justifications: the participants stated that due to their shared characteristics, the robots and the humans should be treated similarly. In contrast, the Three Laws allow treating robots quite objectively, prioritizing the good of humans and neglecting the protection of the robots' existence.

5. Conclusion

The main objective of this paper was to examine people's expectations towards the moral attitudes of social robots. The conclusions are based on the results of three empirical studies in which the stories of robots (and humans) acting in hypothetical scenarios were employed and the moral acceptance of their attitudes was assessed. The similarity in the evaluation of the humans' and robots' morality manifested in respondents' expectations that the robots would act in accordance with certain moral attitudes, just as they would expect from the humans. The question, therefore, was what kind of attitudes do we expect from robots and whether these attitudes should also be identical for both agents?

Each study shows some differences in the moral attitudes expected from the robots and the humans. Malle *et al*. (2015) demonstrated that robots are more strongly expected to make the utilitarian choices (sacrificing one person in order to save four). Laakasuo *et al*. (2019) found expectations of both robot and human's attitudes were strongly related to respect for the patient's autonomy in that the robot nurse's decisions were less acceptable only in the forceful medication condition. Finally, it seems that in certain circumstances the protection of a human's life is considered more morally right than the protection of the robot's being. Differences that emerged in Malle *et al*. (2015) can be partially explained by the context effect: the order in which the stories with the robot and the human agent appeared had an impact on the respondents' judgements. It is possible that in the current study, where stories with the robot agents always preceded those with the human agents, a similar effect occurred. Perhaps presenting the stories in a different order would reveal more differences in the assessment of the human's and the robot's attitudes. This would mean that when people evaluate the attitudes of a robot first, their evaluation is mainly based on their opinion about people and only when they are told to evaluate the attitudes of a human first, differences in their assessments of the two agents appear. Future studies could investigate whether this phenomenon actually occurs.

There were, however, a number of similarities in the assessment of robots' and humans' attitudes. The overall amount of blame received by both a human and a robot agent was similar, which contributes to the claim that the moral decision-making capacity makes the robots natural targets for moral blame (Malle *et al*. 2015). An additional study (Voiklis, Kim, Cusimano, & Malle 2016) analysed the justifications for moral judgements provided by the respondents in Malle *et al*. (2015). It was demonstrated that even if sometimes different moral attitudes were expected from the humans and the robots, participants often provided similar types of justifications for their moral judgments. This suggests that people extend their moral reasoning (or moral intuition) to robots, regardless of the norms applied. In Laakasuo *et al*. (2019), the strong impact of the prospect of losing autonomy resulted in no difference in the evaluation of the robots' and humans' attitudes. Moreover, the decision to respect the patient's autonomy was considered more approvable than forcefully medicating, regardless of the agent.

Laakasuo *et al*. (2019) demonstrated that in certain circumstances people make similar consequentialist moral judgements when evaluating both the human and the robot decisions. However, Malle *et al*. (2015) suggest that in some extreme cases consequentialist moral judgements are made differently depending on the agent being evaluated. According to the current findings, apart from the First Law, Asimov's laws were considered moderately right, both with regard to the robots and to the humans. The incongruity of the First Law stems from harmful situation in which the agent cannot defend itself. Therefore, according to the results, the robots should protect their existence and obey people, but in some situations, they should be able to hurt a human (in self-defence, defence of other people, or other values). Consistent with the above is the result of the MFQ, suggesting that the more progressive the respondents were, the less they thought Asimov's Laws should apply to robots. As the participants' justifications indicate, The Three Laws allow robots to be treated objectively, while the respondents expected them to be treated similarly to human beings. The aforementioned findings could make an important contribution to the discussion of whether robots should have the status of moral patients and moral agents (e.g., Sullins 2006; Hoffmann & Hahn 2019). They are also consistent with the criticism of Asimov's laws in this context (see Anderson 2008).

The fact that Malle *et al*. (2015) and Laakasuo *et al*. (2019) reported more differences in the evaluation of the robots' and the humans' attitudes than the present study may be explained by the character of the experimental task used in these studies. It seems that such an extreme task as the trolley dilemma or the scenario in which the patient is deprived of their autonomy triggers some differences in the moral judgements.

These results could be of considerable use both in implementing morality into robots and in the legal evaluation of their attitudes and behaviour. Malle and Thapa (2017) revealed that the desire for Social-Moral Skills in robots increased over the years 2013-2016. The present work answers the question of which moral skills people expect. An awareness of the strong influence of the prospect of losing autonomy and the need for explanatory skills as well as empathy will improve the designs of nursing robots. In their detailed overview of AI ethical guidelines, Hagendorff (2020) states that most of the guidelines omit contexts of care, nurture, help, welfare, social responsibility, or ecological networks, and so they lack an interpretation of moral problems within a wider framework of "empathic" and "emotion-oriented" ethics of care. As the current findings have shown, this context of understanding the morality of robots is of huge importance to humans. Therefore, taking into account people's expectations can create better AI guidelines.

The fact that the robots are required to make utilitarian choices may prove potentially useful in the context of choices made by autonomous cars, highlighted in the introduction of the present paper. Regarding military robots, people consider their existence to be less valuable than a human soldier's life and believe that robots can be sacrificed in the name of other values. Also, the context effect described in the results of Malle *et al*. (2015) may occur in real life, for example when a legislative body evaluates the behaviour or rights of

a robot by comparing it with those of humans.

Just as Hoffmann and Hahn (2019) recommended people are familiarised with how AI algorithms work, it is also important to take into account people's opinion on what they expect from robots' moral (and any other) attitudes. As noted in Ljungblad *et al*. (2011), robots' ethical concerns should be grounded in the empirical data and not limited to the philosophical considerations. Although the present paper fulfils this purpose, another critical issue highlighted by Ljungblad *et al*. (2011) is that these studies should not be based on futuristic scenarios and robots that do not exist yet. Nevertheless, all the robots and the situations described in this review were hypothetical. Therefore, in order to reveal the ethical implications that may be missed while using speculative scenarios, future work should concentrate on "the actual use of existing robots in a real environment" (Ljungblad *et al*. 2011, 191).

An undoubted weakness of the presented studies is the relatively small number of respondents. The project designed by scientists from MIT Media Lab[1] may be the answer to this problem and thereby constitutes the future of research on ethical issues related to AI. The project aims to collect people's insights into the ethics of robots through crowdsourcing and simple games. The authors state that "The Moral Machine" attracted worldwide attention, and allowed them to collect 39.61 million decisions in 233 countries, dependencies, or territories (Awad *et al*. 2018, 60). Thanks to this method we can examine what decisions people think robots should make when faced with moral dilemmas.

References

Anderson S. 2008. "Asimov's Three Laws of Robotics and Machine Metaethics," *AI & Society* 22:477–493.

Asimov I. 1981. "The Three Laws," *Compute!* 11(18):18.

Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.-F., & Rahwan I. 2018. "The Moral Machine Experiment," *Nature* 563(7729):59–64.

Fong T., Nourbakhsh I., & Dautenhahn K. 2003. "A Survey of Socially Interactive Robots," *Robotics and Autonomous Systems* 42(3/4):143–166.

Giger J.-C., Moura D., Almeida N., & Piçarra N. 2017. "Attitudes Towards Social Robots: The Role of Gender, Belief in Human Nature Uniqueness, Religiousness and Interest in Science Fiction," in S. N. de Jesus & P. Pinto (Eds), *Proceedings of II International Congress on Interdisciplinarity in Social and Human Sciences,* Vol. 11 (p. 509). Research Centre for Spatial and Organizational Dynamics, University of Algarve Faro, Portugal.

1 https://www.media.mit.edu/projects/moral-machine/overview/

Hagendorff T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds and Machines* 30:99–120.

Hoffmann C. H. & Hahn B. 2019. "Decentered Ethics in the Machine Era and Guidance for AI Regulation," *AI & Society* 35:1–10.

Jarmakowski-Kostrzanowski T. & Jarmakowska-Kostrzanowska L. 2016. "Polska adaptacja kwestionariusza kodów moralnych (MFQ-PL)," *Psychologia Społeczna* 11:489–508.

Laakasuo M., Kunnari A., Palomäki J., Rauhala S., Koverola M., Lehtonen N., Halonen J., Repo M., Visala A., & Drosinou M. 2019. "Moral Psychology of Nursing Robots – Humans Dislike Violations of Patient Autonomy But Like Robots Disobeying Orders." URL: https://psyarxiv.com (retrieved on July 18, 2020).

Ljungblad S., Nylander S., & Nørgaard M. 2011 (March). "Beyond Speculative Ethics in HRI? Ethical Considerations and the Relation to Empirical Data," *Proceedings of the 6th International Conference on Human–Robot Interaction (HRI)* (pp. 191–192). IEEE.

Malle B. F., Scheutz M., Arnold T., Voiklis J., & Cusimano C. 2015 (March). "Sacrifice One for the Good of Many? People Apply Different Moral Norms to Human and Robot Agents," *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 117–124). IEEE.

Malle B. F. & Thapa Magar S. 2017 (March). "What Kind of Mind Do I Want in My Robot? Developing a Measure of Desired Mental Capacities in Social Robots," *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 195–196). IEEE.

MoralFoundations.org 2016. "Moral Foundations Theory." URL: https://www.moralfoundations.org/ (retrieved on January 5, 2019).

Murphy R. & Woods D. D. 2009. "Beyond Asimov: The Three Laws of Responsible Robotics," *IEEE Intelligent Systems* 24(4):14–20.

Sullins J. P. 2006. "When Is a Robot a Moral Agent," *Machine Ethics* 6:23–30.

Voiklis J., Kim B., Cusimano C., & Malle B. F. 2016 (August). "Moral Judgments of Human vs Robot Agents," *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 775–780). IEEE.