VILNIUS UNIVERSITY

IEVA VASILIONYTĖ

THE POSSIBILITY OF A MORAL THEORY COMPATIBLE WITH COMMON-SENSE MORALITY

Doctoral dissertation

Humanities, Philosophy (01 H)

Vilnius, 2014

The dissertation was prepared at Vilnius University, years 2008-2014

Research supervisor:

Assoc. Prof. Dr. Nijolė Radavičienė (Vilnius University, Humanities, Philosophy – 01 H)

Consultant:

Prof. Dr. Alvydas Jokubaitis (Vilnius University, Humanities, Philosophy – 01 H)

VILNIAUS UNIVERSITETAS

IEVA VASILIONYTĖ

KAIP GALIMA SU SVEIKO PROTO MORALE SUDERINAMA MORALĖS TEORIJA

Daktaro disertacija

Humanitariniai mokslai, filosofija (01 H)

Vilnius, 2014

Disertacija rengta 2008-2014 metais Vilniaus universitete

Mokslinis vadovas:

doc. dr. Nijolė Radavičienė (Vilniaus universitetas, humanitariniai mokslai, filosofija – 01 H)

Konsultantas:

prof. dr. Alvydas Jokubaitis (Vilniaus universitetas, humanitariniai mokslai, filosofija – 01 H)

Contents

Introduction

The problem. Meta-ethics is often understood as an attempt to understand presuppositions and commitments of moral talk and practice. In the process of constructing a moral theory, the practical character of morality claims its share as moral agents cannot avoid assuming the first person point of view and a common-sense perspective. Therefore, the criticism to the extent that "it is counterintuitive" or "it clashes with common-sense morality" is in many cases fatal to a moral theory. This fact indicates that contemporary meta-ethics aims at embodying presuppositions and commitments of common-sense morality.

But why common sense and common-sense morality? After all, common sense consists of the widespread pre-theoretical convictions, or opinions which seem to be obviously true, and it has been a target of philosophical criticism for centuries as a conglomerate of superstition. However, at the same time many philosophers – at least since Plato and up to nowadays – have seen it as the best place to start the quest for truth. It would have served either as a block of opinions that needed purification from errors and inconsistences or as signposts diverting from far-fetched philosophical speculations.

On the one hand, common-sense moral beliefs are challenged by both the ordinary folk and philosophers and so it is a natural place to start checking their reliability reflectively. It is, after all, what meta-ethics is after: reflection of presuppositions and commitments of moral thought and practice. And, from a methodological point of view, it is as good a starting point as any other, or even better (for explication, if needed, see Appendix 1).

On the other hand, we can ask what kind of moral theory can withstand the blows of human experience, or the criticisms of common sense. It is especially pressing in moral philosophy if moral theory is to be a theory about and for actual human beings.

It is well known that common-sense morality is pluralistic. But, however vast the variety of ordinary moral practices may be, analysing such practices, as well as the main debates in contemporary analytic moral philosophy, one finds that there are two fundamental aspects of moral practices, or two fundamental suppositions of common-sense morality. One of them concerns the truth-aptness and the other – the practical character of moral judgements. To put it otherwise, we talk and act as if our moral judgements were in some sense objectively right or wrong *and* as if *at the same time* they were necessarily action-guiding. Naturally the next question is: if it is so, how can moral judgements have such, on the face of it, incompatible features?

Because if correct moral answers are made so in virtue of a correspondence relation with some kind of objective moral facts, it means that these answers represent the world the way it is. But how can the acknowledgement of facts, of the way the things *are*, be a direct indication of what we should do, of the way the things *should be*? And to the contrary, practical guidelines ("do this", "do not do that") do not seem to be truth-apt, at least not in the same way that factual propositions about the world are truth-apt.

The most popular theoretical positions in meta-ethics exclude either one or the other of the said features: for one part of them, moral judgements describe states of moral affairs, the other part holds them to be imperatives or expressions of, e.g. emotions or pro- and con-attitudes, or acceptance of systems of norms, or acceptance of plans. So the question arises if a moral theory which embodies *both* of our fundamental features of common-sense morality, or our main suppositions of moral practices, is possible at all. And if so, how? In other words, can our common-sense morality, as defined by its main characteristics, be correct? It is answering this question that the dissertation is dedicated to.

Thus, in this work, "common-sense morality" is not a whichever body of opinions on morality, but is rather defined by the two aforementioned

necessary features, i.e. from a common-sense point of view, judgements that lack these two features can be anything but moral judgements.

The thesis and the main claims of the dissertation. I claim that a moral theory which embodies the two fundamental features of common-sense morality is possible, only if it makes coherence its constitutive value and uses the approach of rationalist internalism. This thesis is grounded in the following lines of argumentation:

- From a methodological point of view there are two varieties of moral realism that embody the common-sense approach to moral reality and seek to account for the truth-aptness of moral judgements: the mind-independent (MR_{MI}) and the mind-dependent (MR_{MD}) variety of moral realism.
- Truth-aptness of moral judgements is viably explained only by the MR_{MD}, the position which relies on rationalist epistemology.
- It is rationalist epistemology that allows for an inclusion of the element of practicality of moral judgements into theory, i.e. it is the rationalist construal of internalism (a position defending an essentially action-guiding, or practical, character of moral judgements) that is viable.
- It is the interpretation of rationality as *primarily* coherence that enables the incorporation of both fundamental features of common-sense morality into an adequate moral theory.

Coherence in philosophy is usually understood in its negative sense, i.e. as absence of incoherence (absence of inconsistency or other clashes of beliefs). In this work, however, coherence, following Harman (2002), is conceived also in its positive sense – as consisting in connection of support (such as that of explanation, generalisation, implication or similar) between various states of mind. Besides, the notion of coherence in its positive sense is extended to cover also the relations between propositional attitudes or mental states other than beliefs: it may be a harmonious relation between beliefs and desires or intentions or practical beliefs.

Aims and tasks of the dissertation. In this dissertation I aim, first, at evaluation of the plausibility of the theoretical models of common-sense morality. For that, I will set several different criteria of evaluation and apply all of them to the theories in question.

Second, I seek to analyse, reconstruct (where needed) and reinforce a particular version of the most promising model of common-sense morality, i.e. rationalist internalism. Detailed examination of the accounts of Christine M. Korsgaard and Michael Smith, as well as the main criticisms of their accounts and my own original contributions, will serve the purpose.

Third, I want to reveal the necessary conditions for the incorporation of both fundamental features of common-sense morality into a successful theory of morality. In order to achieve that, I will need to pinpoint the element unique to the successful theory and lacking from its closest rival.

And forth, I want to re-evaluate the most common distinctions in the meta-ethical debates. So I will discuss, question and subtly, but importantly re-define the dominating understanding of the distinctions of realism/anti-realism, cognitivism/non-cognitivism and description/prescription.

Relevance of the dissertation and previous research on the topic. The questions that are analysed in the dissertation, i.e. questions of relations between moral judgements and motivation, or the action-guiding aspect of moral judgements, as well as problems of moral cognitivism and moral realism, or explanation of the truth-aptness of moral judgements, attracts unceasing attention of the academics: a great many articles are being published in such important academic journals as *Ethics*, *Journal of Moral Philosophy*, *Analysis*, *Philosophy and Phenomenological Research*, *Mind* and others. Not to mention a number of anthologies, collections of papers and monographs by numerous philosophers and interdisciplinary researchers, including such celebrated authors like Simon Blackburn (1993, 1998, 2010), Derek Parfit (1984, 2011a, 2011b), Jonathan Dancy (1993, 2000, 2004), Michael Smith (1994, 2004), Christine M. Korsgaard (1996, 2008, 2009), the late Sir

Bernard O. Williams (1981, 1985 and others) and many others, which have been published in the last four decades.

With the recent advent of a new form of research, that is group research financed through project-activities, several philosophical projects have been financed and carried out. To mention but a few: a project *Emerging Themes in 21st Century Meta-Ethics: Evaluative and Normative Language* (2012-2013) at The Edinburgh Centre for Epistemology, Mind and Normativity; a project *Moral Motivation: Evidence and Relevance* (2010-2012) at the University of Gothenburg; several projects (e.g. *Agency and Values* or *Personal Autonomy, Addiction and Mental Disorder*, etc.) at the University of Oslo, Centre for the Study of Mind in Nature.

But the relevance of the present research is witnessed not only by abundant academic interest in the said problems. The question of whether our common-sense understanding of morality – at least as it is defined by its fundamental features – is well or ill-founded, is one of those questions that never lose their importance for non-philosophers as well. In various spheres and situations of life people ask these questions and share their answers, even if not in such a fluent and technical language as that of philosophers. This research contributes to these standing debates and proposes a picture of common-sense morality that is plausible theoretically – it offers such morality solid foundation.

Rationalist internalism in its current guises and thus labelled is relatively new as a moral theory, and its greatest representatives are still developing and refining their theories: Korsgaard has laid foundations to her account in (1996) and has refined it in (2008a, 2008d, 2008e, 2008f, 2008g, 2008h, 2009) and elsewhere; Smith has made a powerful statement in (1994) and has been developing and clarifying his views in (1995, 1996, 1997, 2001, 2002a, 2002b, 2004a, 2004b, 2004b, 2007, 2009) and elsewhere. There are numerous publications regarding the position, only a part of which I made use of in this dissertation: e.g. Gert (2008), Nichols (2002), Mason (2008), Strandberg (2012b), Strandberg and Björklund (2013), Zangwill (2008, 2012). Thus,

rationalist internalism is in the making, and, as many questions remain unresolved, it invites further examination and improvements.

To the best of my knowledge, in Lithuania the person who has tackled a big range of meta-ethical questions extensively is Professor Jūratė Baranova in (2004). In the latter book Professor presents the main moral theories of the XXth century and looks for their relation to the ideas of Immanuel Kant. But while my and Professor‘s enterprises are in several respects parallel to each other, whereas Professor chooses a thorough discussion of a number of authors and relates them directly to Kant, I opt for a more fundamental analysis of only some of meta-ethical positions and for a detailed analysis of the contemporary incarnation of Kantian moral views.

There are several more Lithuanian publications concerning meta-ethics in some way or another, e.g. Patapas (2001), Kuzmickas (1989), Jokubaitis (2013), however, they are not directly relevant to my research. In other words, in Lithuania the contemporary rationalist internalism has, so far, not been given the much deserved attention.

Novelty and significance of the dissertation. The novelty of the present research lies, first of all, in the very project of writing a work of such structure: it aims to show the superiority of rationalist internalism in relation to all other meta-ethical theories trough investigation of the logical possibilities of meta-ethical positions based on the available choices of methodology, epistemology, ontology and semantics.

Besides, I introduce several restrictions and re-define several positions, which either has not been done before or was not brought to its logical conclusions. For example, I re-define moral realism and distinguish between its two varieties; I formulate the proportionality/commensurateness requirement and clearly separate the unconditional and conditional as well as the restricted and unrestricted versions of motivational internalism. Finally, I present some original arguments and analyses in favour of rationalist internalism, such as the

analysis of acting for the sake of the bad or good in the subsection on moral fetishism or the criticisms to the unrestricted motivational internalism.

Methodology of the dissertation. Cognitivism and internalism are the two essential premises which I rely on in the dissertation and which have determined the structure and the extent of the present research.

First, I hold that cognitivism (the view that moral judgements are truth-apt) is the dominant semantic position in meta-ethics, therefore, I do not examine the non-cognitivist theories. Validity of this supposition is supported by the results of a survey conducted by Bourget and Chalmers (2013), a detailed analysis of which can be found in Appendix 2.

Given that cognitivism about moral judgements is the "received wisdom" and one of the fundamental suppositions of moral practices, I concentrate on the two possible explanations in virtue of what moral judgements can have truth values, i.e. on the analyses of the two (the mind-independent, MR_{MI}, and the mind-dependent, MR_{MD}) varieties of moral realism. After giving the reasons to accept one of them rather than the other, I turn to internalism (the view that moral judgements are necessarily action-guiding, expressive of the second fundamental feature of common-sense morality) in connection to cognitivism.

Second, I limit my attention to internalism, because I hold that the need of externalism in connection to cognitivism is usually determined by acceptance of the mind-independent version of moral realism, i.e. by one's choice to account for the truth values of moral claims by their correspondence to the state of moral affairs. Once the latter is ruled out as the best explanation of cognitivism, the reason for choosing externalism is usually gone with it. So in Part II I discuss externalist arguments only in as much as they target the main points of rationalist internalism, and the rest of this part of the work is dedicated to the explication, reconstruction, interpretation and reinforcement of the latter position.

In Part I, I mainly concentrate on the "negative" defence of the position that finally will prove to be able to embody the two features of common-sense morality (i.e. MR_{MD}): I point out the multiple flaws of its closest rival (MR_{MI}). There I expose one of the great controversies in meta-ethics, thus the grain of analysis is rather coarse. Meanwhile, in Part II, I present a detailed explication of a particular version of the mind-dependent moral realism – rationalist internalism, thus the fine-grained analysis. In general, the motivational internalism/externalism debate is very technical and specialised – as confirmed by the aforementioned survey results (to be found in Appendix 2).

Appendix 1 is an explication of how meta-ethics is to be conceived and how this conception dictates the goal of the present research, and it includes a presentation of the possible methodological approaches in meta-ethics, as well as a justification of my choice of the moral realist methodology.

Appendix 2 contains my analysis of the results of a survey by David Bourget and David J. Chalmers, which supports my choice of structure and of different grain of analysis in the two parts of the dissertation.

The two appendices are useful for locating the theories and discussions of this research in a wider context of meta-ethics.

Structure of the dissertation. I will begin the enterprise with showing which methodological approach in meta-ethics is preferable for our purposes in Chapter 1 of Part I. Given that (methodological) moral realism is such an approach, Part I will be dedicated to the analysis of its two versions – one of which defends a view that the truth making conditions of moral judgements are mind-independent, and the other one – that they are instead mind-dependent. I will refer to them, accordingly, as the mind-independent and the mind-dependent varieties of moral realism.

Having provided the standards for an adequate meta-ethical theory in Chapter 1, I will be exposing flaws of the mind-independent variety of moral realism throughout the whole of the Chapter 2. In the end of Part I, the question of normativity (touched upon in 2.1.) will be revisited which will

serve as a bridge between the two parts: we finish discussion of failures of one variety of moral realism with a promise that the other one will be able to cope with the challenge of normativity, and that promise is kept in the next part.

In Part II, Chapter 1, I introduce the most general conception of motivational internalism which embodies the feature of practicality of the moral judgements. It is shown that in order for it to be a plausible claim several refinements are to be introduced, as well as some terminological questions to be settled. The most promising refined version of motivational internalism, that is, rationalist internalism, becomes my focus in Chapter 2. I analyse extensively the conception of rationality which is at the core of the latter position in Chapter 3. Chapter 4 analyses the possible psychological models of rationalist internalism and Chapter 5 – the related topics of autonomy and normativity. In Chapter 6, I finalise the research with a short discussion of the relation and value of rationality and morality.

Part I

Accommodating common-sense morality: truth-aptness of moral judgements

1. Moral realism and criteria for an adequate meta-ethical theory

The preferred methodology. Meta-ethical enterprise is often (and is in this research) just conceptual, meaning that meta-ethicists seek to make sense of suppositions, however, it is an open question if a certain theoretical picture of morality refers[1] to anything actually (we can only present some inductive arguments in support of such hopes). In general, it is possible to approach the question of how to build (or test) a body of knowledge, or to construct a theory yielding knowledge about the world (moral or otherwise) that we live in in several different ways. For example, Roderick Chisholm (1977/1966 and 2001/1973) discerns three such ways due to the logical possibilities to answer the two most general questions of epistemology: *What* do we know?" and "How are we to decide, in any particular case, *whether* we know?" (Chisholm 1977: 120).

Chisholm claims that in order to answer one of these questions we are required to answer the other one, so we are necessarily caught in a vicious circle: in order to know if things are really the way they seem to be, we must have a procedure for distinguishing the true appearances from the false ones, but in order to know if our procedure is good, if it succeeds in distinguishing them, we should know which appearances are true and which false (Chisholm 2001: 190). Chisholm calls this "the problem of the criterion".

One of the possible views with regard to this problem is scepticism which takes the gravity of the problem to block the possibility of any solid solution.

[1] [1] That is, if something is conceptually possible, it is not necessarily ontologically possible in our world. And, surely, if something is possible in our world, it does not mean that it is also actual, but there being no actuality of a certain moral order as presupposed by some theory is not as crucial as the impossibility of such a moral order for undermining the relevance of the moral theory which presupposed it.

Another possibility, called by Chisholm particularism, is to answer the first question of the extent of knowledge and, based on that, to answer the second one (of the criteria of knowledge); the third possibility, called methodism, is to begin with the second question and to proceed to the first.

On the one hand, choice of the methodological approach is arbitrary and there is no non-question begging reason to favour one starting point over the other (that is, once you question your opponent's position, you assume one of the other two positions). On the other hand, in view of the goals of the present dissertation there are several reasons to favour one of the approaches over the others (see Appendix 1), and one reason is especially weighty. We are up to finding a moral theory which is compatible with the common-sense morality (as defined by its two fundamental features), and I claim that such a theory has to embody the value of coherence. On the methodological level this means that people's moral knowledge should be coherent with their moral practices, i.e. the actual functioning of morality should not be different from our knowledge of its functioning.

Particularism, which can also be called common-sensism, is an optimistic, or even a naïve position: it is based on trust that our moral practices are basically on the right track, that people can discern the main aspects of moral reality and so that in their main beliefs (as to the character of morality) they do not err. It allows people to have access to that reality without any specific tools, without being privileged. So particularism purports to give a transparent theory, i.e. such that the true nature of the requirements of morality would be accessible to the ones subject to it, thus preserving the integrity and autonomy of the moral agents.

Meanwhile, scepticism puts a person into a strange position or a strange state of mind: one has to act on what very well may be or even is a mistaken knowledge. I call it "schizophrenia" in its etymological sense of "split mind": a person believes one thing, but acts on another, and – what is more – by her/his own lights. Such a split is rather likely to be obtained also by the theories based on methodist approach, because they are likely to produce a very restricted and

in many aspects counterintuitive view of reality which conflicts with some of the fundamental aspects of common-sense understanding of reality. Theories that separate the truth of the theory from the truth of the practice, threaten the effectiveness or even autonomy of the agents and make ethics a subject of political agenda (what behaviour is it best that people stick to?) or a subject of science.

That is why I choose to investigate only those moral theories which embody the particularist approach which consists in acknowledging that we do know certain ethical facts, or in acknowledging some moral phenomena the status of reality based on common sense. This naïve methodological approach can also be called common-sensism due to the fact that this position gives credit to a common-sense view of the world, or in virtue of the importance it bestows on common sense at the beginning of the theoretical quest. It can equally well be termed “realism", or “(methodological) moral realism” in case of meta-ethics.

However, I am well aware that “moral realism” is a problematic label. Nowadays, it can be attached to positions ranging from Moorean robust moral realism often associated with Platonism to those moderate ones which are simply adverse to relativism. But when I used this term in the aforementioned sense, I meant it as a methodological position. In this sense realism is a position which begins the quest for knowledge from assigning some of the phenomena the status of reality or verisimilitude. It is in this sense that I understand moral realism in this work, and we will see shortly that it is possible, and even preferable, to do so from the perspective of several other philosophers as well.

On the moral realist approach, after deciding on the extent of moral knowledge, one then proceeds to answering the question of what epistemic pathways lead us to the moral knowledge, thus, moral realism can be realised in different ways – depending on which of the source(s) of moral knowledge one chooses to defend.

Fundamental features of morality. We can ask what features are to be considered fundamental, how to discern those salient common-sense features of *moral* reality. It seems an easier thing to do in non-moral phenomenology, where resilience of reality is more palpable and the non-constructivist nature of the reality behind those phenomena (at least for most theorists) is apparent. It is more difficult with morality. Still, we can say that some of the features being given up, the talk of morality would lose its sense (for example, we could talk of etiquette instead) and its practices would not be *moral* practices any more. Those are the constitutive features of morality. A good way to unearth them is not by explicitly asking people what features they consider to be constitutive of morality, but by looking at what silent (pre)suppositions their moral practices are based upon, i.e. by examining which practices and expectations are default – common and automatized.

In the moral realm there are two suppositions that are essential, i.e. two features that meet criteria for constitutive features of moral reality: the cognitivist and to some extent objectivist and the practical character of moral judgements. In other words, a supposition that morality is objective (not an expression of one's preferences or desires – unless accidentally so[2]) and that one is necessarily motivated by what one judges is the right thing to do (moral motivation is not contingent upon the character traits or accidental desires that a person may or may not have at some moment of time).

I should emphasise that the two suppositions ground more than just moral practices of ours, so one should not be surprised that our talk will often swing from "moral judgements" and "moral practices" to "practical judgements" or "normative judgements", or that we will engage in comparisons of theoretical and practical thought, theoretical and practical reasoning. But I have to say that while the analysis of moral judgements depends on a more general analysis of practical judgements, I leave it open which other (than moral) kinds of judgements enter this category (i.e. if aesthetic judgements are such

[2] I underline again the basis for eliminating forms of cognitivist moral relativism from the present research.

practical/normative judgements or not, etc.). Morality surely has its specifics, but I will talk about it later on.

The importance of the two suppositions is confirmed by the fact that they are also the main target of the meta-ethical theories: the two out of three main debates are the cognitivist/non-cognitivist debate and the internalism/externalism controversy (the third is that of realism/anti-realism). Theorists defend or try to explain away at least one of the two features.

It is also confirmed by our practices. Let us put it in short, and then elaborate. The cognitivist character of moral judgements is presupposed by our practices of moral arguments (at least of the meaningful ones): it only makes sense to argue if there are correct (and incorrect) answers to be had to (at least the main) moral questions and that by giving each other reasons for some position or other we stand a chance of obtaining such answers. The practical character of moral judgements is presupposed by our expectations that people act in accordance with what they sincerely judge to be the right thing for their own selves to do. That is, the belief is that people not just talk in vain, but that moral answers matter practically: people are (at least usually or at least under certain conditions) necessarily and not by chance motivated in accordance with their own moral judgements.

The common sense theorists begin with our common-sense assumptions that ground our practices, such as practices of conversations[3]. The fact that we often bother conversing with others, that we take clashes in beliefs to signal a need for clarification of the reasons for our differing views, thus, that we take the contrary beliefs of others to constitute a challenge to ours, shows that this practice of conversation relies on certain premises concerning the correct formation of beliefs and the abilities of our conversational partners. Premises, as Smith puts it, about "the norms to which the believers … are subject, and about the capacities they enjoy" (Smith 2004a: 85), that is, that there are

[3] Structure and functioning of ordinary language as embodiment of common sense is an important object of investigation for theorists of this kind: Moore, Reid, Smith, Williams (in 2006/1985) – all recognise its value.

certain norms that govern belief, that believers are capable of recognising those norms and that they are capable of responding appropriately to that recognition. If we did not grant the believers those certain capacities, then discussing matters and trying to get people to believe things through conversation would be futile. Thus, these are the suppositions our conversing practice relies on. And, according to Smith, the same goes for intrapersonal conversations which are nothing else than thinking; that way "To call into question the propriety of making these assumptions is thus to call into question the propriety not just of conversing with others, but of all thought" (Smith 2004a: 89). Naturally, the same goes for *practical* interpersonal conversations, where people are treated as potential *agents* rather than *believers* in the narrow sense.

Two points have to be stressed. First, such an approach does not suppose that people *always use* these capacities. On the one hand, "people can retain their capacity to recognise and respond to the norms that govern their beliefs even when they fail to recognise and respond to those norms on some particular occasion" (Smith 2004a: 88). On the other hand, "there are various conditions believers can be in that remove – whether temporarily or permanently, locally, or globally – their capacity either to recognise the norms that govern their beliefs, or their capacity to adjust their beliefs in response to their recognition of such norms, or both. Unconsciousness, illness, stubbornness, arrogance, self-deception, and drunkenness are some among them" (Smith 2004a: 88).

However, these two aforementioned fundamental features of morality seem to pull into opposite directions, to be incompatible. If moral judgements are truth-apt, how can they be practical (no *ought* from *is*)? If they are action-guiding, how can they be truth-apt? This difficulty to combine them into a coherent moral theory, taken at face value, divides philosophers into two groups: those defending the cognitivist character of moral judgements and those defending the practicality of moral judgements (as *the* defining feature of moral judgements). As Smith puts it, by pulling against each other, these

features threaten "to make the very idea of morality altogether incoherent" (Smith 1994: 5). Thus, the task of the philosopher who adopts the common-sensist strategy is clear: "to make sense of a practice having these features", "two of the more distinctive features of morality, features that are manifest in ordinary moral practice as it is engaged in by ordinary folk" (Smith 1994: 4-5). So there is also a third way – to deal with the difficulty, i.e. to stick to the thought that an adequate moral theory should incorporate both features and propose such a theory.

What hinges on the (im)possibility of a moral theory that incorporates both features. As mentioned above, without these features moral arguments would lose their point: either moral judgements would lose their authority and become a matter of taste or otherwise subjective attitudes, or the making of moral judgements would have no reliable relation to our actual motivation, "failure" to comply to one's own normative judgements would not indicate anything at all, i.e. moral judgements would have no practical implications and there would be no difference between cases of what we now call "weak-will" and the so-called "normal" cases. Moreover, as Smith and Pettit note, were people mistaken in postulating freedom of thought and action and were they to embrace this knowledge, "They would have to discount everything they must assume in order to practice conversation, and relate more broadly in an interpersonal fashion ... in order to think" (Pettit and Smith 1996: 447).

For some of the theorists who try to reconcile the two features, the inner coherence of persons (not to have to separate the truth of the theory from the truth of the practice), as well as meaningfulness of our thoughts and actions[4] is extremely important, hence the task of proposing such a theory that would be in harmony with the practice. To "make sense" is one of the keywords of these

[4] Indeed, an action for such theorists just is a unit of meaning, not a combination of bodily movements and – what is even more important – not a product of mere *post factum* rationalisation. It is rather the correspondence of the contents and of quantitative characteristics of the states of mind which are constitutive of action.

theorists who take our human condition seriously. One can quote Frankfurt to show the underlying motivation of this strand of philosophy:

> "Taking ourselves seriously means that we are not prepared to accept ourselves just as we come. We want our thoughts, our feelings, our choices, and our behavior to make sense. We are not satisfied to think that our ideas are formed haphazardly, or that our actions are driven by transient and opaque impulses or by mindless decisions. We need to direct ourselves—or at any rate to believe that we are directing ourselves—in thoughtful conformity to stable and appropriate norms" (Frankfurt 2006: 2).

What hinges on the possibility of a moral theory incorporating both basic features, is not only a preserved sense of meaning, but also the authority of morality. The problem of the authority of morality is mainly related to the cognitivist character of moral judgements, but not limited to it. If there is no truth to be found about morality, if it is a matter of taste or expression of personal preference (not subject to reasoned change), why should it be authoritative or any more authoritative than any other inner tug? But if we can find the moral truths out, still, why should they be authoritative with relation to our behaviour any more than any other kinds of truth about the world? And how could those truths be necessarily action-guiding?

So the same problem of the compatibility of the two features can be seen as the problem of authority of morality and the task of a philosopher, starting from a belief that morality matters, is then to show why morality deserves to be our practical guide, how it earns its credentials so we can let it lead our way. Let us remember that such a question could well arise for any person, not just the professional philosophers, whenever one's beliefs or practices get challenged. But a philosopher's answer will be more technical. In this case her task is to give such an analysis of moral judgement that would show how a moral judgement can be both truth-apt and have a practical upshot, or how practical knowledge is possible.

There is one more problem. If the internalist cognitivist picture of moral judgements cannot be correct, then moral judgements have no more intrinsic authority than any other kind of judgements (e.g. aesthetic judgements, requirements of etiquette or driving rules – depending on the opposing views). But the impossibility to differentiate between moral judgements and other kind of directives are characteristic of psychopaths. And there being no reliable direct relation between our judgements expressive of our values and our motivation eradicates a difference between the reason-based decisions and a pattern of fixated motivation, such as that of a fetishist.

In other words, if the current cognitivist and internalist assumptions cannot be put into a coherent moral theory or if they do not actually obtain, then there is no possibility to distinguish a psyche of a fetishist or a psychopath from the psyche of a supposedly normal person in the moral sphere. For example, Smith claims that when trying to account for a seemingly reliable relation between moral judgements and respective motivation, externalists posit a certain desire to be moral, and that just turns morally good people into moral fetishists. However, if morally good people are reliably motivated to do what they believe they should do (and not because of direct care for others and their causes), there is no significant difference in their motivation and the motivation of psychopaths who refrain from something only because or do something despite "it's not the done thing"[5].

Authority of morality as a criterion of adequacy. The task of the meta-ethical theories, or the criterion of their adequacy, can also be formulated in terms of the "authority of morality". We perceive morality as authoritative, and authoritative in a special sense. This supposition underlies our moral practices, our moral judgements. According to empirical tests, psychopaths do not see any difference between moral authority and authority of conventions such as driving without license or playing with one's food, etc. (Nichols 2002).

[5] I take the data on psychopaths from Nichols (2002). I analyse this data and other aspects of a psychopaths' understanding of morality in more detail further on.

Children from a young age, as well as psychologically normal adults, including criminals, make "a significant moral/conventional distinction on permissibility, seriousness, and authority contingence" (Nichols 2002: 14). That is, analysing the empirical testing data, from normal subjects' answers one could work out that it is less permissible and more serious to make moral transgressions (with relation to conventional transgressions) not just because of their social unacceptability, but because of the unfairness to the victim[6] (ibid.: 13-14).

Such an authority of morality is independent on any other specific authority (such as that of other people with power to punish or so, or on God). To refer to a rephrased answer to a Euthyphro dilemma, it is not because God says it is bad that immoral actions are bad. Surely, people may disagree on which norms are conventional and which truly or strictly moral (in the sense of the norms that persist through time and space, independent of the passing moods or changing customs), that is, about the extent of morality or about the contents of it. However, what matters here is the very fact that we make this distinction and that we make it in virtue of the mentioned features which define morality (along the other-regarding character of it). And I call this a presupposition because it is not necessarily reflected upon and it is not what the tested people said they thought about the authority of morality, but what could be deduced from their differing judgments about the cases presented to them.

An adequate meta-ethical theory should be able to keep this authoritativeness of morality and explain it. Authority of morality can be explained *away* by explaining the *seeming* authority in *non-moral* terms. However, I said I would be interested in the non-sceptical positions in this work. The question of authority of morality is sometimes called "the normative question": why should morality bind us, why should we be subject to it, or simply – why be moral? In other words, the quest for normativity of morality is

[6] Whereas psychopaths "were much less likely than the control criminals to justify rules with reference to the victim's welfare. Rather, psychopaths typically gave conventional-type justifications for all transgressions (e.g., "it's not the done thing" [the subjects were British])" (Nichols 2002: 14).

the quest for its sources or for the grounds of its authority. Therefore, an aspiration of a successful meta-ethical (not a psychological, sociological or other) theory is to answer the normative question, to unveil the sources of moral normativity.

But what is it that people are doubting when they preoccupy themselves with the question "why should *I* be moral?". I believe there can be several worries behind this. In some cases one doubts moral requirement to be a fair requirement, that is, a just requirement, in other cases one wishes to make a moral judgement into one's own decision, to meaningfully relate to it.

In one sense "why should I" may be a very personal question: why should it be *me* who does it. After all, morality does not stop demanding you to save a drowning child just because there are other people nearby. One way of answering the worry is by showing that it addresses everyone or anyone. People feel that equal treatment of everyone (who is equal to others in relevant respects) is part of the idea of justice, and so it is just to be required what everyone else is required to do. However, even if people feel victims of injustice if someone else is exempted from some requirement, they often feel comfortable if the exception is granted to their own selves. They also know that systems usually do not get destabilised because of one exception or two: perhaps if it is only me that does not obey, the system will not crumble. That is a well-known problem of a "selfish knave" or (more neutrally) of a "free-rider"[7].

But if authority of morality is dependent on the authority of society (or a care for its well-being) or on the overall social outcomes of people's individual behaviour, such an authority is not enough to sustain a persuasive "you should". Then, a normative question rightly expresses doubt of whether moral requirements do not just cover up an interest of somebody (group, say, society; or individual, say, a king or a prime minister whose interest is to "keep them

[7] The danger of such thinking is the more apparent the more people succumb to it. A guise of this problem, I believe, is also the so-called bystander effect researched in social psychology.

all in line"). In asking such a question one needs to make sure that one is not being deceived or manipulated. One asks why somebody's interest is more important than one's own.

But how can morality be objectively/intersubjectively valid and not to cover an interest of somebody else[8]? The question may be understood as requiring a convincing answer that, despite appearances, it somehow is to my own benefit, it is in my interest to act morally.

So one possible answer may be given in terms of interests. In many cases we do not just have one and only possible way of action, but we choose from two or more of them. There are obvious benefits (immediate or not) to be achieved or own (or of the ones we care about for some reason) interests served in many of those cases, but not always in cases of moral behaviour. In this respect a Kantian understanding of the relation between happiness and morality is more in line with contemporary thinking than the one of the ancient Greek philosophers: morality does not necessarily lead to happiness, though (perhaps) it makes you worthy of it[9]. In other words, no immediate interest for

[8] Morality is quite obviously to the benefit of the other one, but the other one who? There is a difference between my action serving another for gaining power or other goods, and my action benefitting another person as a human being – despite one's particular goals. It is more soothing to think that I, as a human being, owe another human being decent behaviour, which makes both of us into human beings, rather than that I should sacrifice my particular interests of my well-being for somebody else's particular "worldly" interests. The interest of the person in need of my moral action is not covered, it is apparent, so in the text above I go on to explore questions and answers concerning other interests than those of the subject of one's moral action.

[9] Think of the folk understanding that good people do not deserve bad things happen to them. If bad things happen, they often look for a reason: if bad things happen to good people, folk thinks it not only unfortunate, but also unjust; if it happens to bad people, the folk often thinks of it as of a punishment, retribution or "a lesson", sent by the fate, by universal justice or by God (or maybe accidental, but leaving one with a feeling of deserved justice). So the idea that moral qualities make you worthy or unworthy of happiness is there because it is right, or just, to deserve good things happen to you.

In other words, people are more skeptical of the sufficiency of good character for happiness. Greek ethics is rather egoistic in the sense that it is concerned with perfection of the character of a subject, which makes the strife for good life, or happiness, or flourishing, coincide with the strife for self-perfection and so depend to a great extent on the person himself. One can be virtuous and happy – at least to a large extent – *despite* the unfortunate circumstances and despite others. Especially

the self may be served. One can remember that moral phenomenology is such that the beneficiary of a moral action is exactly the other, not the self. In that case, the question amounts to asking how it is to my interest to choose an action that is not obviously in my interest, the more so – why should I prefer it to other actions, embodying more immediate interests of mine?

So in a sense, one wants to know if moral requirements are just in the sense that they do not cover foreign interests and do not make me (or us) into an object of manipulation. The claims of morality have to be objective in a sense. At the same time, knowing that they do not represent anyone's interests would not make them authoritative. Requirements should represent such interests of mine that deserve my reverence. So this "objectivity" should be such as to represent the interests of each and every of us, but not of anyone in particular (or not so particular that I could not identify myself with).

This question (of the normative basis of morality) is usually asked or at least is especially pressing, when "the faith wavers", or, as Korsgaard would put it, when it requires us to do something hard[10]. It means that in such minutes

---

think of Platonic contemplation of ideas and of stoic passionless person (but not of Aristotle).

Meanwhile in the Modern times, ethics is concentrated on making the human relations, their co-existence and interactions agreeable. The well-being becomes dependent on both related sides. And ethics is put in terms of duties, or obligations: everyone has to contribute to justice.

Whether we deal here with a secularised version of the idea of desert for one's actions based on their moral character, or not, the worthiness to be happy seems to spread the requirements of making the kind person happy onto other people and onto the circumstances, leaving the kind person just partly in control of his/her happiness (or at least of the share of the happiness that is connected to morality).

However, one should bear in mind that a virtuous person is – at least usually – much more than a morally good person, and so that contemporary moral philosophy is – usually - much more restricted in scope than the ancient Greek ethics. Moral philosophy is not concerned with a personal flourishing, because a person's flourishing is due to so many things in human life that cannot be reduced just to the contentment of being a moral person. One can surely notice that morality – in so far as it is concerned with the relations of people – contributes to the flourishing of personal relations, but relation is always dependent on at least two people, besides, there is more to personal fulfilment and well-being than good relations with others (as important a part as they can be).

[10] I.e. "when what morality commands, obliges, or recommends is *hard*: that we share decisions with people whose intelligence or integrity don't inspire our confidence;

we doubt the very importance of our moral interests. The answer, accordingly, should be persuasive to such a person under pressure.

The normative question can be put in a slightly different wording still. The normative question does not just ask for an explanation of morality and its authority, but also for its justification. An adequate moral theory should harmonise both functions – that of explanation and justification. A moral theory cannot be proper if one can understand how morality came about and admit that particular moral claims mean exactly what some moral theory says it does, but still not to see how the explanation guarantees the possibility of justification. Smith claims: "Someone who says 'Though it would be right to act in that way, there is no justification at all for doing it' mis-uses the word 'right'" (Smith 2004c: 202). I.e. if a theory explains the authority of morality so that in the eyes of the one whom it is explained it loses the authority once had, the explanation apparently came apart from justification. Korsgaard also argues for the normative or justificatory adequacy of a theory of moral concepts (Korsgaard 1996: 13). And so I also agree with the requirement that an adequate theory must meet this criterion.

To sum up the criteria for a proper meta-ethical theory: an adequate meta-ethical theory should contain both suppositions – a cognitivist and an internalist – or, to put it otherwise, to be able to preserve the normativity/authoritativeness of morality, or, again, to both explain and justify moral claims. I stress that all three differently formulated criteria will allow for or exclude the same theories.

Moral Realism(s). One way to understand moral realism is to define it as consisting of such two claims: "(1) moral … claims are capable of being true or false; and (2) some of these claims are true" (Street 2010: 370). In other words, moral realism consists of a cognitivist position (1) which opposes non-cognitivism (moral claims are not truth-apt; thus, expressivism, emotivism,

---

that we assume grave responsibilities to which we feel inadequate; that we sacrifice our lives, or voluntarily relinquish what makes them sweet" (Korsgaard 1996: 9).

prescriptivism and the like), and the claim that sometimes truth making conditions of the moral claims obtain (2), which is contrary to moral nihilism. Thus, cognitivism is necessary for moral realism, but not sufficient for it. For example, error-theory, most famously advocated by Mackie[11], is a combination of cognitivism and nihilism: it accepts (1) and rejects (2).

It is in this sense that Smith uses the term of moral realism when referring even to what other theorists would call constructivism (Smith 2004c). And it is precisely the meaning of Korsgaard's term "procedural realism" in (Korsgaard 1996)[12].

Thus understood, moral realism includes a wide range of views – including constructivism, and even, according to Street, "a simple subjectivism according to which what's good for a person is whatever that person thinks is good" (Street 2010: 370). On this definition of realism, it is a position that does not specify *in virtue of what* the moral claims are true or false, what their truth making conditions are. It only expresses approval of the idea of available ethical knowledge (whatever sense we put into "knowledge").

Two notes are to be made here about the claims (1) and (2), though. First of all, "true" should not be understood in a truth minimalist sense. As Street notices, drawing on minimalist theories of truth, such expressivists as Simon Blackburn and Allan Gibbard would agree with the claims (1) and (2), so these claims are commonly restated adding a qualification "in a non-minimal sense" or "in a strict sense" (Street *forthcoming*: 40, n. 8). However, the same effect as

[11] Nowadays the error-theory is still defended by R. Joyce, R. Garner A. Miller and others.

[12] Korsgaard brings out the ambiguity of the meaning of "realism" in contrasting what she calls "procedural realism" with "substantive realism". On her view, in its minimal sense, realism, or procedural realism, is a logical opposite to skepticism, to nihilism. It is a position that there are correct answers to moral questions, right and wrong ways to answer them, whereas its opposite denies the existence of moral truth (Korsgaard 1996: 34-5).
In 2008f/2003, though, she says that "Moral realism, rather, is the view about *why* propositions employing moral concepts may have truth values" (2008f: 302), whereas on the question of truth values realists and constructivists can agree. This meaning of "moral realism", I gather, expresses the same idea as that behind her former term of "substantive realism".

adding these qualifications can be achieved by demonstrating that minimalism about truth is inadequate as a theory of truth, as, for example, Smith did in (Smith 2004c) and in (Jackson, Oppy, and Smith 1994). Let me summarise the argument of (Smith 2004c) in several sentences.

Minimalism, according to Smith, fails to explain "what it is about a sentence that is capable of truth and falsehood that makes it capable of truth and falsehood" by claiming it is a purely syntactic feature of the sentences (Smith 2004c: 185-186). Minimalists say that the strings of, say, English words are truth-apt if they are "of an appropriate grammatical type" to figure in a whole array of contexts: as the antecedents of conditionals, in propositional attitude contexts and so on (ibid.). However, Smith shows that even the nonsense sentences (such as from Lewis Carroll's *Jabberwocky*) meet these criteria without being meaningful and so without being truth-apt. Therefore, he claims that the idea of mere syntax being sufficient to establish truth-aptitude is absurd. Regardless of whether one finds Smith's argument successful or not, one should keep in mind that the definition of realism above is only valid with a non-minimalist conception of truth.

Another point to make is about the distinction of cognitivism/non-cognitivism. Formulation of the cognitivist claim (i.e. (1): "moral … claims are capable of being true or false") is rather wide and neutral, even if we agreed that it excluded minimalist reading of "true". It does not presuppose a specific conception of truth, nor a specific psychological position, that is, whether the claims are true because they correctly describe the facts or on other grounds, or whether those claims express "ordinary beliefs". So in view of (1) such philosophers as Korsgaard are to be considered as cognitivists, even if she would not approve any of the latter specified positions.

If it was otherwise, the cognitivist/non-cognitivist distinction and division of the specific philosophers into respective "camps" could be called to question. For example, Street notes that the distinction or, to be more precise, one possible understanding of it under which the claim is not as neutral has been called to question by both Gibbard who is an expressivist, thus, a non-

cognitivist, and Korsgaard who is a constructivist, thus, a cognitivist. The grounds for doubt would be such: 1) on the psychological level, none of them think that normative, including moral, predicates are used to express states of mind which are ordinary beliefs (Street 2010: 376); 2) on linguistic level, they do not believe that normative claims, or predicates, *describe reality*. From expressivists' point of view, such predicates even do not describe the mental states of the people uttering them, but *express* those mental states. From constructivists' point of view, the role of such predicates may be defined in the following way.

According to Korsgaard, for constructivists, normative concepts, first of all, are the names of solutions to practical problems: a normative concept refers to "whatever solves the problem", and the conception behind that concept proposes a particular solution. She gives an example from Rawls's *A Theory of Justice*: there is a distribution problem in a society, and the concept of justice names a solution to that problem, whereas the conception of justice is "a principle that is proposed as a solution to the distribution problem" (Korsgaard 2008f: 322). Therefore, we get truth when a concept will be applied correctly, whereas its correct application is guided by a correct conception, and a correct conception is that which solves the problem, not that which describes correctly some mind-independent reality. However, it can describe reality, but that human reality which is constructed (ibid.). In other words, normative property concepts do not *denote* natural properties; if normative claims describe something, it is those states of affairs (actual or not) that conform to a respective normative principle.

Thus, the distinction between cognitivism and non-cognitivism should not be understood in terms of difference in views on the question if moral claims can be true or false in virtue of describing moral reality (*vs.* prescribing, expressing something) and on whether they express a cognitive state of mind (as in (theoretical) belief *vs.* desire). Such a distinction would leave out "theories like Aristotle's and Kant's, according to which moral judgements are

the conclusions of practical reasoning", that is, neither obvious descriptions of facts about the world, nor emotional expletives (Korsgaard 2008f: 309).

Hare, for example, argues for a hybrid character of the moral statements (they share characteristics of descriptions and prescriptions), thus, he calls attention to the same problem: accounts that do not qualify as genuinely descriptivist or prescriptivist fall through the cracks. He seems to advocate a view that prescriptions need the descriptive meaning to explain them, so "an initial dogmatic insistence" that the moral statements are descriptions block the explanation of the moral statements as hybrid. Moral utterances, Hare notes, have a rather firm descriptive meaning to them, so that we know what non-moral properties substantiate those moral claims. However, it is a different question if some or other society is right to assign moral terms this or that descriptive meaning, or, in other words, if it is right or wrong to recommend or condemn certain kinds of acts.

And this discussion of the possible readings of the cognitivist claim brings us to the further question of truth conditions. It is with a further restriction on the nature of the truth making conditions of moral claims that a more restricted meaning of "moral realism" is obtained. The difference between such newly obtained more robust realism and the rest of the positions that also accept the claims (1) and (2) can be spelled out in terms of *mind-dependence*: are the truth making conditions mind-dependent or rather mind-independent? For a *robust* moral realist, what makes moral claims true are the features – natural or not – of the mind-independent world. As Korsgaard puts it, moral realism in this sense is "the view that propositions employing moral concepts may have truth values because moral concepts describe or refer to normative entities or facts that exist independently of those concepts themselves" (Korsgaard 2008f: 302). The opposite position is usually termed "anti-realism" and holds those truth making conditions to be mind-dependent. The latter position harbours constructivism and forms of moral subjectivism.

Coming back to Korsgaard's distinction between "procedural realism" and "substantive realism", the difference can also be put in the following way:

"The procedural moral realist thinks that there are answers to moral questions *because* there are correct procedures for arriving at them. But the substantive moral realist thinks that there are correct procedures for answering moral questions *because* there are moral truths or facts which exist independently of those procedures, and which those procedures track" (Korsgaard 1996: 36-37).

Thus, moral realism in its robust sense involves an additional layer of metaphysical commitment to moral entities or moral facts, to weaving morality into the fabric of the world (to use a phrase coined by Hare). For a moral realist of this kind, ethics is part or continuation of non-moral epistemology, a theoretical enterprise, even if the nature of moral properties possibly differs from the nature of the non-moral properties. Indeed, for a moral realist the possible distinction of epistemology and moral epistemology does not exist: "For the moral realist, ethics and metaphysics are not separate areas of philosophical inquiry. To be a moral realist is to take a position on what the world is like" (Jackson 1998: 204). In this more robust sense "moral realist" is not just a methodological position anymore, it is its variety, incarnation or realisation which can be called "metaphysical moral realism" or "mind-independent variety" of the methodological moral realism (MR_{MI}) as opposed to the "mind-dependent moral realism" (MR_{MD}) which I will prefer to "anti-realism".

If considered to be a form of anti-realism (according to the wide-spread aforementioned distinction), constructivism, along with moral subjectivism, has to meet the doubts about the objectivity of morality. This denomination is especially unfair to constructivism which has to prove that even if moral reality exists only *in relation to* human beings, still, contrary to relativism, this normative reality *does not depend on* human beings: there still are moral facts to be found out by human beings, or the right answers to moral questions, and those answers are not relative to the changing and contingent features of people's background, but they are rather determined in relation to the necessary features of the human psycho-physical make-up. So I will rather

prefer treating constructivism as the second incarnation of the methodological moral realism and use indexing to distinguish between the two varieties.

However, I want to note that when exploring the mind-dependent variety of moral realism, I will be talking about constructivism rather than forms of moral subjectivism akin to relativism. Moral relativism has many faults, the greatest of which is adversity to common-sense understanding of morality in many respects. To mention just a few: it contradicts the idea that morality is somewhat objective, that moral arguments are meaningful (but there is no real argument if two people just ascribe themselves different attitudes), it dissolves the idea of moral truth into a banal and uninformative reports of genuineness of psychological states. Because of these features, the goal of the dissertation lets us leave moral subjectivist theories out of the present discussion.

When talking about moral realism in the more robust meaning I will use the MR_{MI}. Knowing that its non-naturalist version (e.g. Moorean realism) has received too many criticisms to be still as popular nowadays as at the beginning of the last century, I will have the naturalist – whether in its reductionist or non-reductionist guise – moral realism in mind[13].

2. Disadvantages of the mind-independent moral realism

2.1. The insurmountable distance between human interests and the good

Moral realism and construal of objectivity. So according to realists, morality is objective, but for them more than one conception of objectivity is available. For moral $realists_{MI}$ moral knowledge is as objective as knowledge

[13] Naturalist moral realism is a position that moral properties, entities or facts are explicably related to (respectively, reducible to or irreducibly supervening on) natural properties, entities or facts. I define what I mean by "naturalist realism" at the same time being aware that this term can sometimes have a different meaning: it can be used to refer to the idea that moral entities are natural in the sense of having causal powers – just like natural particles or objects, such as stones etc. It is in this sense that Street assigns the label of "non-naturalist realism" to Ronald Dworkin, Thomas Nagel and similar philosophers (Street *forthcoming*, esp. see 13-14). However, I will use "naturalist" in a loose sense that will become clearer further on.

of the natural world can be, because, as mentioned, for them ethics is continuous with metaphysics.

There can be several reconstructions of realists' motivation for construing objectivity in such a way, but whatever the motivation, the objectivity of morality will result from correspondence of moral language with moral facts or truths. As Putnam notes, a philosophical idea since Plato is such that "if a claim is objectively true, then there have to be *objects* to which the claim 'corresponds' – an idea which is built into the very etymology of the word 'objective'" (Putnam 2004: 52).

Korsgaard believes that the motivation behind realism$_{MI}$ in general is "the sense of impeding loss" (Korsgaard 1996: 47). According to her research in the history of modern moral philosophy, moral realism$_{MI}$ always arises as a response to a threat posed by somebody ("a self-proclaimed spokesperson for the Modern Scientific World View", ibid.) who challenges the possibility of moral knowledge. If the challenge is that the ethical knowledge is impossible, a realist will try to prove that it is possible. However, the problem here, according to Korsgaard, is that nobody asks if ethical knowledge needs to be defended at all, or, to put it otherwise, if the challenge is worth accepting[14]. It seems that the question, spelled out in a specific way, programs the answer, as the modern realists$_{MI}$ accept the sceptics' criteria for a good answer by sharing the same "Modern Scientific World View". "And so long as moral realism appears to be the only alternative to these skeptical options [relativism, scepticism, subjectivism], the need to show that moral truth is as solid, as real, as objective, as scientific truth – will seem pressing" (Korsgaard 2008f: 309).

Blackburn has similar ideas: he thinks that what is threatened is the power of morality to obligate. His thought is that people feel unease because of the tension between the subjective source of morality and its objective feel, or phenomenology. If morality is not objective, it would not have "the power or force, the title to respect" which we think morality does have. Blackburn believes that obligation needs to be perceived as "something sufficiently

[14] Accepting along with the criteria for knowledge.

external to us to act as a *constraint* or bound on our other sentiments and desires", that it "must come from outside of us" (Blackburn 1985: 6). In other words, in order for the "must" not to be conditional upon our desires, we should feel that the requirements are at least partly external to us, that there must be a principled possibility of distance between the norm and un-normed directives. Thus the defence of the objectivity and of an external source of obligation/morality.

Hare also gives his explanation of why some people become moral realists$_{MI}$: it is because of their belief that unless wrongness is part of the fabric of the world, there will be no way of rationally determining if some act is wrong. And that is due, according to him, to a prejudice about rationality represented best by Hume's views on the functions of reason and conception of rationality. To give the citation: "Reason is the discovery of truth and falsehood. Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason (*Treatise*, III, 1, i)" (Hume *apud* Hare 1985: 48). Whether this genealogy of moral realism$_{MI}$ is correct or not, Hare is right about the widespread relation of moral realism$_{MI}$ and Humeanism in moral psychology, i.e. a narrow understanding of rationality (rationality can only be instrumental) dependent on a narrow understanding of the functions of reason[15].

Meanwhile, Hare claims that Humean conception of rationality is a prejudice as apparently we can rationally decide what to do or what to ask or what to advise others to do. He highlights that thought processes which have prescriptions as their end products, can be rational, and their rationality does not depend only on the rationality of the fact finding process (Hare 1985: 49).

[15] The latter idea is defended in (Korsgaard 1986). And, as already mentioned, Korsgaard traces the reasons for defending moral realism$_{MI}$ to the same fears of the impossibility to obtain moral knowledge.

Another way to think about how contemporary moral realism$_{MI}$ is born is to think about its task of explaining how morality is possible in the natural world, or the task of showing how moral reality is compatible with the modern worldview. However, this task also reveals that the criteria for that which is real are set so that moral reality (and moral knowledge) needs to fit the requirements for natural reality (and natural knowledge).

The need of such a robust sense of objectivity (and of reality or existence, and of knowledge accordingly) has twofold implications: 1) on a conceptual level, one should treat moral claims as descriptions; 2) on metaphysical level, one needs a plausible ontology because of the need of the correspondence relation between the language and reality: "if you regard some value judgments as objectively true, you will conclude that they are *descriptions*; and if you cannot construe them to your own satisfaction as descriptions of natural objects and properties, you will be forced to construe them as descriptions which refer to non-natural entities" (Putnam 2004: 52-53). As to the latter point, it is apparent that in such a case moral objects or properties will need to be analysed reductively (or, less plausibly, non-reductively) either in natural terms or in non-natural terms (a rather implausible choice nowadays).

However, it seems that realism$_{MI}$ can at least preserve the authority of morality: if such an objectivity of moral knowledge is achievable, no one can suspect it to cover up the interests of somebody. But as we will see, such a model neither preserves the unconditional authority of morality, nor gets the character of ethical inquiries right.

Objectivity of morality as alienating people's interest in morality. The problem is that construing objective as *neutral* to specific interests that anybody can have, means that it is even not (necessarily) to the interest of the one who makes the judgement. As we very well know, scientific truth does not by itself imply any behavioural directive. To elaborate on an example from Zangwill (2012), from the fact that Mount Everest is 8,848 meters high, or from the fact that it is the highest mountain on the Earth, or from the fact that it

is part of the natural world, it does not follow neither that I should climb it, nor that I should like it, nor anything else of the kind. An answer that George Mallory, a now late English mountaineer, "is reputed to have given to the question of why he wanted to climb Mount" – "Because it's there" (Zangwill 2012) – is neither a typical one, nor a satisfactory one. That is because the applications of the knowledge about natural world depend on the aims, or interests, that people have. People's interests do not alter the character of knowledge/truth about natural world, and this knowledge/truth does not dictate the ultimate aims of people. The relevance of such kind of knowledge/truth is a function of people's interest.

In other words, to think of that which morality requires as equally alien to everyone's interests, is not an answer. It must still be possible to hold the obligation to be your own in some way, it must not come *totally* from without: "It is as if the objectivists' error is to think of certain things as obligatory in a way which has nothing to do with us, and about which we can do nothing: a way which could in principle stand opposed to the whole world of human desire and need" (Blackburn 1985: 7).

Then, the vision of ethics as a subject and of its uses is the following. Ethical life is a matter of application of ethical knowledge: "The moral realist thinks of practical philosophy as an essentially theoretical subject. Its business is to find, or anyway to argue that we can find, some sort of ethical knowledge that we can apply in action" (Korsgaard 2008f: 325). Several things follow. The relevance of moral knowledge, or moral truth, becomes dependent on every person's aims, or interests, as moral truth does not imply any behavioural directive.

In other words, no *ought* from *is* – Hume's law is not broken. So far so good, but that means that even if the authority of morality does not depend on the covert interests of other people, the authority of morality becomes conditional upon a specific interest of the person who makes a specific moral judgement, i.e. either on extra-moral interests or on the moral interests a person

acts on. If for the former, morality has no authority of its own, if for the latter, we will see that for such persons the charge of "moral fetishism" is in place.

To sum up the problem, as Zangwill nicely puts it: "even if morality is there in the world, it is not clear why we should concern ourselves with it, any more than we ought to be concerned with methane in the rings of Saturn" (Zangwill 2012: 345). By the way, Hume's name did not come up accidentally: because of this structural peculiarity of moral realism$_{MI}$, Humean model of moral psychology becomes acceptable to the realists[16] (at least to those who want to avoid positing prescriptivity built into the moral entities or properties, thus, to avoid a suspicious ontology). However, the normative question (Why should I care about morality? Why should I necessarily apply the moral knowledge?) is not answered.

An externalist answer to the normative question. Some realists would certainly disagree that it is not answered. For example, Zangwill takes the normativity challenge seriously, but ends up with the claim that the only *justification* of authority of morality just is *metaphysical* ("Because it's there"). To wish for a different kind of answer is to beg the question. For him, it is true that the relevance of moral knowledge to a person depends on that person's desires. However, the *justification* of the authority of morality *is not to be found from the first-personal point of view*, it is not to be found in our mind (Zangwill 2012: 360).

I have to draw attention to the understanding of "justification" here: what justifies is not, for example, good enough arguments which are better than those to the contrary, but the existence of moral facts which may be unknown to the person oneself. Justification for Zangwill and many other moral realists$_{MI}$ is detached from rationality, which is apparent from his understanding of justification of morality and from his criticisms of the Kantian position: "this view is unKantian in that it opens a sizable gap between

[16] That is, beliefs are motivationally inert, so in order to explain an action, one needs to have a pair of a means-ends belief and a desire.

moral and rational norms" (Zangwill 2012: 360)[17]. Accordingly, reasons should be understood as existing apart from human mind – be it actual or idealised (!).

Meanwhile, for realists$_{MD}$, as one could expect, justificatory reasons are mind dependent. It does not mean that each reason that a person from one's own point of view thinks to be a (good) reason, is such indeed. One can be wrong. The standard for a justificatory reason under certain defined circumstances is set by an idealised human mind. However, this position admits both that people act without being perfectly informed and that we know that we have been wrong only in face of better reasons and better arguments (the idealised human mind is getting more perfect and errs less in time faced with new evidence and better arguments). So according to the MR$_{MD}$, we should not despair that our actions are perhaps never justified objectively in the most robust sense of objectivity: we are justified when acting to the best of our knowledge rationally and being open for rational discussions of our reasons, because it is us who create and perfect them (even if their standards are not up to our whims or wishful thinking).

"Authority" in MR$_{MI}$ gets a specific meaning as well: it is not something people voluntarily acknowledge to something or somebody in virtue of their valuable properties, but something in the nature of facts, knowledge or truth, i.e. if you know about that thing's existence, the fact of its existence should be authoritative to you. In case of MR$_{MI}$, though, this authority is not necessarily motivating.

It is a typical realist$_{MI}$ position. Zangwill, however, tries to replace the comparison of a moral fact to the Mount Everest with a metaphor of the world as a moral minefield. Zangwill's idea is that as we can only instantiate natural properties, and some of these are "attached" (through supervenience relation) to certain moral properties, so we thereby instantiate moral properties. Thus, moral properties "snare" us. Supposedly, "that is the only possible explanation

[17] One can put it in terms of internalist/externalist views in epistemology (see Zangwill 2012: 351).

of why we should heed moral demands. We should heed them on pain of instantiating negative moral properties. The natural world is a moral minefield! Step on the wrong natural property and a negative moral property explodes. We must be careful where we tread" (Zangwill 2012: 362).

But as is often with metaphors, they are impressive till untangled. Whereas the powerful image of losing one's limbs or even life when strolling through a minefield suggests big losses to one's own precious physical self, the moral minefield does not have the same effect. In morality the one who gets hurt is primarily the other and in many cases the damage for one's self is not direct and not as damaging or fatal. To remember, quite a big part of morality is concerned not with the well-being of the agent[18] (understood as individualistically as it usually is) and it hardly should be. The latter is surely important, but to think that the effects of bad behaviour are primarily damaging the one who performs it, means presupposing the agent is morally sensitive and gives importance to morality in the first place, i.e. that one cares about having clear conscience and thinks bad actions have disfiguring effect on one's character *beside* the direct care for others.

So the metaphor is not as good suggesting also that the motivation should be connected to avoiding either damage to one's self or just the moral badness happening in the world. As Korsgaard has nicely put it in her writings, moral $\text{realists}_{\text{MI}}$ are right *in a sense*: "Mackie is wrong and realism is right. … For it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. They are people, and the other animals" (Korsgaard 1996: 166). However, their mistake is essential: to detach our direct objects of care from morality thus leaving morality somewhere on the Mount Everest which nobody knows why one should climb.

In any case, the answer to "Why should I not instantiate a negative moral property?" does not come as directly and obviously as the answer to "Why

[18] To remember Nichols's inquiry and the normal people's view that the impermissibility of moral transgressions is related to the unfairness to the victims of immoral behaviour.

should I not take a walk in a minefield?". And the answer to the former is not "I will suffer". What Zangwill, along with other realists can offer as an answer here is "the only normative question is one that is answered by the existence of moral facts. It is just part of being a moral realist that one refuses to answer the normative question beyond a certain point" (Zangwill 2012: 347). And that is just that – refusing to answer the normative question and stamping one's foot to one's initial claim.

Certainly, one can remain on the realist$_{MI}$ side claiming that the normative question does not allow for a certain theory to be fit. As Zangwill notices, how one conceives of a normative question, varies with one's metaphysics (Zangwill 2012: 347). If that is so and if, as he thinks, there is no neutral normative question that could be used as a basis for assessing metaphysical theories of morality, then perhaps we should evaluate the plausibility of the metaphysical and epistemological theories in their own right (and so we will shortly).

Moral fetishism. Usually moral realists$_{MI}$ cannot answer why one should act on moral beliefs more plausibly than in terms of instrumental rationality. When considering the question of the value of true beliefs to a person, Zangwill claims true beliefs not to have any obvious intrinsic value, but only instrumental value as they enable us to satisfy our desires (Zangwill 2012: 351-352). So ultimately, according to him, it is the desire-satisfaction that is the source of relevance of moral beliefs and of the value of rationality: if we did not care about anything, we would not need to be rational[19], but we do care about at least one thing. Such views make the value and so the normativity (defined, as Zangwill would say, not from MR$_{MI}$ perspective) of moral truths/knowledge dependent on a person's desires. Which only brings us to one possible explanation of the reliable relation between moral judgements and moral motivation, which could satisfy those who ask for a *moral* source of

[19] I want to underline the understanding of the value of rationality according to this position, as we will see a very different alternative to this understanding.

motivation for moral actions: to posit a desire to be moral. Respectively, a desire to be moral calls for application of the moral knowledge. And quite some realists$_{MI}$ take this way.

For example, Svavarsdóttir claims this desire to be moral to amount to "wanting to take the morally justified option" (Svavarsdóttir 2006: 177) thus connecting it to moral reasons and justification (as the desire itself cannot do the justificatory work): "those who are motivated by moral considerations will see these considerations as presenting a claim on them to act or live in a certain way or, at least, as presenting a moral justification for such an action or a way of living" (ibid.: 177). Another philosopher, Brink, talks about a "desire or other practical commitment to being moral" (Brink 1997: 14). What connects such authors, is, of course, a Humean psychological model of motivation rooted in what Korsgaard calls "skepticism about practical reason": a belief, or a cognitive state of mind more generally, cannot motivate alone, there is a need of a desire, or, more generally, of a conative state of mind, for that[20]. But this is only natural having in mind that the non-practical knowledge (into which they turn the ethical knowledge) is motivationally inert, or insufficient for motivation. At this point, I can again underline that MR$_{MI}$ downplays the role and value of rationality, of reason as ability and of the cognitive states, and this constitutes a sharp contrast with the other alternative realisation of MR[21].

However, being moral for no rationally/intrinsically justifying reason or wishing to avoid harm to one's self is as suspicious a motivation as a desire to be moral (that is, acting for the sake of being moral). So it is this model of motivation based on the desire to be moral that is called by Smith "moral fetishism":

[20] E.g. "What I maintain is that such an acknowledgment of normative standards does not suffice for moral motivation" (Svavarsdóttir 2006: 177) and elsewhere. "If I did not have this more ultimate desire or commitment, my moral belief would lead nowhere (or elsewhere)" (Brink 1997: 14) and "When the virtuous person adds to her background psychological states beliefs about what these moral categories require in particular circumstances, she may be motivated to act, but this will be in virtue of her cognitive and conative background and not simply because of her newly acquired cognitive states" (Brink 1997: 15).

[21] It becomes clear that one alternative is empiricist and the other rationalist.

> "Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read *de dicto* and not *de re*. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue" (Smith 1994: 75).

While arguing against the model of positing a moral desire, Smith invokes Williams (1981b), because the latter advances a similar argument when targeting those moral philosophers who (over)emphasise impartiality. Williams tells a story about a man who faces the choice of saving either his own wife or a stranger, chooses the wife and his motivation is "that it was his wife, and that in situations of this kind it is permissible to save one's wife" (Williams 1981b: 118). Williams claims that such moral philosophers provide the husband with "one thought too many". Similarly, Smith opposes the motivational externalist claiming that in taking the good person to be motivated to do whatever is right, where this is read *de dicto*, they provide the morally good person with one thought too many so alienating her/him from her/his aims, from the direct concern which should instead be essential for a morally good person (Smith 1994: 75-76).

Svavarsdóttir rejects the fetishism charge as "entirely unfounded", because she rather understands moral fetishism as treating morality as sacrosanct: "Moral fetishism is most appropriately thought of as the phenomenon of holding oneself and others to rigorous moral standards, while being completely unwilling to entertain any reflective question about their nature or ground" (Svavarsdóttir 2006: 169).

However, I want to argue that Svavarsdóttir here is wrong. Even if we agreed with her that her account and similar ones do not deserve the label of moral fetishism, it does deserve the charge regardless of what would be the best name for it. One can acknowledge that it is not easy to discern the aims of the activities in the moral case: is it for the sake of the good or for the sake of our more direct care for things that we (should) act? It seems, though, that the

symmetrical case of acting to bring about evil could give us a better understanding of what is involved in the case of acting in order to bring about the good/the right.

For example, in fiction we sometimes meet characters that represent evil in its purest form and who are obsessed with bringing about chaos, destruction and prevalence of the bad to the world. They act in the name of evil and for the sake of it, seemingly just for the fun of it – not for the money or power or anything of the kind. In real life, however, such cases are not pervasive, if existent at all[22]. Even in fiction, not to talk of the reality, most of the baddies do not do something *because* it is evil, but because it is the most efficient way to get what they want: power, revenge (bringing about "justice" of a kind) or similar goods.

Turning to more mundane situations, a notorious historical figure of the XXth century, for example, has done a lot of evil, but not *because* he thought it to be evil, but because he had a very different understanding of what was the right thing to do, or about the well-being or the desirable universal social order for the human beings. So in some of the cases, the bad ones are doing what they do seeking to bring about the good, but their values differ from those of ours. In such a case their understanding of morality may differ from ours rather sharply.

Similar to these cases are those in which people act morally badly when seeking something they value the most, even if *we* think that (at least in these particular circumstances/sphere of activity or in general) this value should not be at the top of the value hierarchy. For example, if people seek profit or power when it clashes with morality or other highly appreciated human values. We think those people act badly, but they seek the things they think are the best – whether they see our point or not. That is, whether their moral views are like ours or not, their value hierarchy holds at its top a controversial good.

[22] I owe thanks to Fritz-Anton Fritzson (Lund University) for bringing this contrast in the case of evil characters in life and fiction to my attention.

In other cases, bad actions are also chosen not *because* they are thought to be evil/bad, but *despite* they are evil or bad. In such cases people who act badly have the same moral norms, but the morally good alternative is downplayed in value – probably temporarily, probably not willingly. For example, the temptation to do something bad, but pleasant is too great to resist, so they do it – not for the sake of the bad, but for the sake of the pleasant, even if they understand that, e.g. adultery is (morally) bad and it rightfully makes them feel guilty.

In other similar cases, like when somebody seeks revenge or hurting somebody, it is also not the evil as such that they seek. It may be the justice in the sense of "a tooth-for-a-tooth" that they are after even if they know they will subject the targeted person to a more or less deserved pain, humiliation or something like it. In these cases they may know that it is bad or not, but it is not for the badness of it that they choose the action. For example, think of a child who felt hurt by one's parents (when they only taught him/her a valuable lesson) and did something nasty knowing it would hurt them; or think of a lover who is seeking revenge.

There may be similar examples, but probably this is enough to prove the point similar to that which Smith is making and that I want to reinforce. In cases when somebody acts in ways that bring about the good, respectively, people do the things they do not *because* they are *good* or *right*, but because of what *constitutes* that goodness. Let us look for a clearer explanation of this claim in Korsgaard's account, especially as the same accusation of moral fetishism usually is advanced to Kantian theories.

Korsgaard believes that many criticisms of the idea of acting from duty are based on confusion and tries to rectify the understanding of what Kant, or at least Kantian ethics, requires: "The idea that acting from duty is something cold, impersonal, or even egoistic is based on the thought that the agent's purpose or aim is 'in order to do my duty' rather than 'in order to help my friend' or 'in order to save my country' or whatever it might be. But that is just

wrong" (Korsgaard 2008a: 218). She does this relying on a distinction between act and action.

According to Korsgaard, both Kant and Aristotle hold an action to be an act-for-the-sake-of-a-certain-end. And it is exactly this whole structure, an action that is an object of evaluation (good/bad). Korsgaard gives an example from Kant: it is not the making a lying promise that is wrong, but making a lying promise (act) for the sake of personal gain (end), and it is not committing suicide, but committing suicide (act) in order to escape your troubles (end) that is wrong (Korsgaard 2008a: 218). It is not that people choose acts, that is, *the means* to the ends which "are foisted upon us by natural forces (e.g. desires)". Humans, according to her, choose the whole package, and a Kantian maxim, as well as the Aristotelian *logos*, targets these duos: only an action can be required by duty, permissible, or forbidden, noble, ignoble or base – not an act.

Alternatively, one can say that what we evaluate is the relation of an end and an act. To say that an *action* is worth doing for its own sake does not mean that an *act* has no *end*. Korsgaard says that morally good *actions* are always chosen for the same reason: because they are intrinsically good, that is, for Kant – because they embody the very form of law, for Aristotle – they are virtuous because they embody the right reason (Korsgaard 2008d: 190-191). However, when we ask why somebody did something, we are rather asking for the explication of the action: it is clear that the agent thought it worth doing, so what is the purpose/end to act relation? Korsgaard notes that usually we know the act, so we ask to reveal the purpose of the act, and we evaluate their relation. She nicely shows by an example that evaluating this relation is not just evaluating if the end is successfully served by the act/means. The example is about a Jack who goes to Chicago to buy paperclips. Though it is true that Jack will be able to buy paperclips in Chicago, but going all the way from Indianapolis to Chicago just to buy a box of paperclips is not worthwhile. It does not make sense.

So coming back to the charge of fetishism, it is a mistake to make a desire to do the right thing into one's purpose of the act, or into one's object of

direct care. The purpose or the object of direct care is "to serve one's country", "to help one's parents", "to restitute justice" and so on. And a good/right thing to do is "to restitute justice by helping to find the culprit" or to "serve one's country by contributing part of one's spare time to help one's fellowmen", whereas a bad/wrong thing to do is, for example, "to restitute justice by murdering the thief that robbed you" or "to serve one's country by spying on every of its citizens". According to the externalist action model, it is constituted by a desire and a means-end belief, but the motivation to do the right thing is embodied by the desire which is part of the action, whereas in a Kantian model the rightness would be a function of the relation of such a belief and desire.

Svavarsdóttir and the like put the cart before the horse. It seems weird to think that persons have (or should have) this disposition to do whatever is right, or the desire to do whatever is moral or justified, and whatever fills this conative state with content, gives them directions to act. As if we were machines waiting for the input so that this reaction between the information that we get and the desires that we have would occur and we would move forward. It should be the other way round. We find ourselves in various situations in life and we have to deal with them. In these situations we get various impulses from our background, such as attachments, duties, wishes and so on (because we never are blank slates), and we ask which of those possible ways of resolving those situations are the right ones: which of those ends and by which of those means we should enact. So I conclude that the realists$_{MI}$ who choose to posit a moral desire are indeed subject to the charge of moral fetishism, and, moreover, that their motivational model implicates a weird view of how people function.

2.2. Inexplicable faith as the basis for moral ontology and epistemology

Moral ontology hangs on faith. The relation of metaphysics and ethics. And so a realist$_{MI}$ moral theory does not meet the internalism criterion (moral judgement does not necessarily motivate to act accordingly). It cannot answer the normative question (or it does answer it only metaphysically) nor to preserve the categorical character of the authority of morality (it is conditional on a person's actual conative attitudes). Zangwill claims that the justificatory work is done by moral metaphysics, but that is a question. Another related question is whether a moral ontology is needed at all, especially as it seems that there is no plausible epistemology that connects to it, so, in other words, it seems that the moral ontology$_{MI}$ only hangs on a belief in it.

An alternative to moral ontology that constructivists, or MR$_{MD}$, propose is that of idealised human psychology. This choice has the advantages of providing a reliable link between the moral realm and human psychology and at the same time preserving the distance between the actual states of human mind and that which constitutes true knowledge of what is moral (preserves the possibility to be mistaken, not to know and, as we will see later, the distance needed for the moral normativity to emerge).

Several theorists think ontology and ethics are distinct disciplines and rightly so. In this respect they differ from realists$_{MI}$ who see ethics as continuation of epistemology and ontology (and believe in the project of ethics as a scientific discipline or as a theory). These adversaries think that ethical theory is a different kind of theory from others, that it rightly belongs to the sphere of practical philosophy and that shifting the perspective from practical to theoretical (or posing scientistic criteria for moral reality) one changes the question and the specific subject matter gets hidden from one.

One of such theorists of the latter kind is Hilary Putnam who approves of Levinas's attitude which is "that all attempts to reduce ethics to a theory of being, or to base ethics upon a theory of being, upon ontology … are disastrous

failures" (Putnam 2004: 23-24). He also joins Heidegger in thinking that "philosophy needs to take the ways of thinking that are indispensable in everyday life much more seriously than the ontotheological tradition has been willing to do" (Putnam 2004: 16). And he is certainly not the only one thinking along the lines: Korsgaard, Smith, Rosati, Street and others sharply criticise $realism_{MI}$ for downplaying the importance of the practical point of view by seeking to ultimately reduce it into the theoretical one and, having failed at that, turning morality into a matter of faith.

As Korsgaard has demonstrated in her (1996), the view that moral, or, more widely, normative, entities or properties or reasons exist rests on confidence that they do[23]. She analyses works of Prichard, Samuel Clarke, Thomas Nagel and others, and concludes that they just find this existence self-evident[24] – that is where the urge to defend a certain position within meta-ethics arises from in the first place. For Prichard and Clarke, for example, the belief that we really have obligations supports the belief in the existence of normative entities or properties. But it simply does not follow. They only get caught in a circle of faith.

According to Korsgaard, Nagel thinks that in order to determine that moral truths or reasons exist it is enough to do the negative job – to rebut scepticism about morality. Once again, $realism_{MI}$ about these things does not follow straight away. Here one can very clearly see that $realism_{MI}$ understands oneself to be the only alternative to scepticism, but that is a false dichotomy. As I have pointed out before, there can be more than one theory that embodies belief in moral reality.

---

[23] Remember, I have placed the MR_{MI} among the methodological realist theories which begin from holding they (normative entities or properties or reasons or truths) do exist.

[24] "These things are so notoriously plain and self-evident, that nothing but the extremist stupidity of mind, corruption of manners, or perverseness of spirit, can possibly make any man entertain the least doubt concerning them" (Clarke *apud* Korsgaard 1996: 39) and "In arguing for this claim, I am somewhat handicapped by the fact that I find it self-evident" (Nagel *apud* Korsgaard 1996: 41).

From Korsgaard's point of view, realism$_{MI}$ cannot answer the normative question because it arises when such our confidence is shaken. Besides, another important question follows: "If confidence can support a metaphysics which in turn is supposed to support the claims of morality, why can't confidence support the claims of morality more directly?" (Korsgaard 1996: 48). This is a very relevant question bringing out the futility of moral ontology. Why posit the entities which are either weird or cause theorists to go into much intellectual equilibristics, if the task of justification of our moral beliefs and requirements can be achieved more efficiently without positing these entities (or properties or whatever it is). The Occam's razor is handy here, but we know this tool must be handled with care, thus, we should not go too far and cut off the moral sphere as such: even if we can explain morality in non-moral terms, we cannot justify it in these.

Street and the moral epistemology of MR$_{MI}$ and MR$_{MD}$. A similar point about MR$_{MI}$ clinging onto nothing else than belief (or faith) is interestingly advocated by Sharon Street. While Korsgaard says that beginning with the obvious reality of moral norms does not lead to positing moral entities, Street concentrates on the aspect of mind-independency of the moral reality that a realist$_{MI}$ advocates. She asks: if this mind-independent realm of morality exists, how should one know that the moral norms we approve of or the moral reasons we have are indeed necessarily representative of that realm?

Street's point can be seen as criticism of the realist$_{MI}$ confidence that the mind-independent reality is correctly represented by moral judgements – even by those that are barely controversial and so inspire our confidence in the existence of the mind-independent moral realm. Street claims that such a confidence is not based on anything, in other words, the answer that we have just been lucky to have landed on the right beliefs is not proper because there is no non question-begging reason for such a belief. But I will present her argument.

Street targets MR_{MI} in the person of Ronald Dworkin in her "Objectivity and Truth: You'd Better Rethink It" (*forthcoming*). First of all, she formulates a puzzle which needs explanation (and the latter, she says, may also be rather simple): "Insofar as we regard our normative judgments as true, we must agree that there is a striking coincidence between (1) the normative judgments that are true, and (2) the normative judgments that causal forces led us to believe" (Street *forthcoming*: 10). The more so, Dworkin himself notices this puzzle in need of explanation.

The puzzle, according to Street, arises because of people's capacity to occupy two points of view – a practical and a theoretical one. The terms she chooses here sound Kantian, even if Street herself is not a Kantian: the claim to the existence of the two perspectives does not seem contentious. To make it totally clear, the practical/theoretical perspectives do embody a Kantian distinction between a point of view from which a person sees oneself as the source of causality, as an agent, and that from which a person sees oneself or other people as part of the world of cause and effect. Street puts the distinction which supposedly is the source of the puzzle thus: when we occupy the practical standpoint, "we understand ourselves as beings who are capable of recognizing what practical reasons we have, what we should or ought to do … and so on. We think of ourselves, in other words, as beings whose normative reasons are true" (Street *forthcoming*: 1). But from the theoretical standpoint we see our normative judgements as "subject to causal explanation", such as "upbringing, cultural background, and inherited psychological tendencies; … had some or all of these factors been different, I wouldn't have made the same set of normative judgments that I now make" (ibid.).

This puzzle is easily resolved by a version of the MR_{MD}, e.g. constructivism, as for the MR_{MD} normative reasons and, accordingly, moral truths are mind-dependent. Were we different creatures, we would most probably have different problems and different solutions to those problems. According to constructivists, the practical problems arise first of all in virtue of a certain structure of mind that we, humans, have. For example, Korsgaard

claims that it is because human beings are reflective and therefore not automatically guided by their instincts that they get the question "what to do?" to arise. Many animals, it seems, just do not face this problem. Therefore, the normative questions, moral ones including, have a very direct link to our natural causally shaped mind, to our history – both personal and that of our species.

The answers are also linked to the mind. Depending on a version of the MR_{MD} theory, the story about this dependency and the strength of it may differ. The more subjectivist the theory, the stronger the link. If we talk about a subjectivist theory (still, of a cognitivist type) which claims that the truth conditions of moral judgements just are the preferences of the agent, then clearly there is no mystery as to the link of the truth of normative judgements and the normative judgements that causal forces led us to believe. However, constructivism is not a subjectivist theory, or if we look from the $realist_{MI}$ point of view, not of extreme subjectivist variety. The answers to the practical problems as understood by constructivists are not relative or arbitrary, they are shaped the way the problems are shaped – by the way the agents are. In a sense, "practical problems … provide standard for their own solutions" as a solution is "arrived at by reflecting on the nature of the problem itself" (Korsgaard 2008f: 325, n. 49 and 322).

Korsgaard gives a good analogy to moral concepts and conceptions (in 2008f: 323): concepts and conceptions of artefacts. She says we have a concept of a chair because the physical construction of humans makes it possible and sometimes necessary to sit down. The concept which represents a solution to a human problem has a fitting conception which gives the solution content. Different cultures have different versions of it, but something is a chair in virtue of the features it has to solve the given problem. The concept of a chair is not arbitrary in that sense. To go on with the analogy, one can see that if moral problems arise in virtue of a certain psycho-physical construction of human beings, answers to them do so as well. Perhaps different human cultures have different versions of them, but it is in virtue of their features which solve

the problem that they are *moral* conceptions. Another task of a constructivist is to make clear in virtue of which part(s) of human construction the moral problems and their solutions arise, but they do it and we will talk about it further on in more detail.

However, for a realist$_{MI}$ the puzzle is a genuine one as for her/him moral truths are mind-independent and s/he must face it. One possible answer is rejected by both Street and Dworkin – that of special moral particles with causal powers (eloquently called by Dworkin "morons"[25]). In this respect, Dworkin and Nagel (also mentioned by Street) agree with Mackie that normative properties or reasons do not play any role in causal explanations. And even if Nagel's thought that "it begs the question to assume that this sort of explanatory necessity is the test of reality for values" (Nagel 1986: 144) would encounter opposition of Sturgeon (for example, in 2006/1985), let us leave his possible objection aside for it not to lead us astray (Sturgeon's views are far from being mainstream in the camp of moral realism anyway).

Dworkin's own answer is that the relation between the causal story and the normative story is that of coincidence, of good luck. There are two different explanations to be had which are not normatively connected: one explanation is normative ("why normative judgments *X*, *Y*, and *Z* are the true ones") and another is causal ("why causal forces led me to affirm those very same judgments"). Street thinks that this may do when talking about an individual normative judgement, but not about our normative judgements in general. She proposes that there are countlessly many internally consistent evaluative systems, so if such internally coherent agents with different moral views lack no non-normative information and make no logical or instrumental errors, what reason does a realist have to think "that the causal forces landed him … on the robustly independent normative truth he posits" (Street *forthcoming*: 20).

[25] Apparently the term is formed in analogy with "proton", "neutron" and such to refer to a moral (subatomic?) particle.

Dworkin admits that in such a case all that a realist can say is those others "did not 'see' or show sufficient 'sensitivity' to what we 'see' or 'sense,' and these metaphors may have nothing behind them but the bare and unsubstantiated conviction that our capacity for moral judgment functions better than theirs did" (Dworkin 1996: 121-122). And that, as Street says, means that a realist$_{MI}$ is in no better position than a person who insists on having (fairly) won the New York Lottery based only on the fact that she entered it – their epistemic situation is just the same. So finally, Street concludes that realism$_{MI}$ in this respect is just "an article of faith", "a strange form of religion" (Street *forthcoming*: 23).

So whereas there is an explanation of conformity between the practical and theoretical standpoints from a constructivist point of view, there is none from a realist$_{MI}$ point of view – unless one has faith there is an inexplicable gift of moral insight granted to the "elected". Street claims that such a faith apart, one must conclude that we are hopeless at discovering normative truth. But such conclusion is totally unfitting as it would paralyse normative reasoning or lead to incoherencies: "I should do *Y*, but I'm in all likelihood hopeless at recognizing what I should do" (Street *forthcoming*: 35). It is as paralysing as scepticism and in the same way – by excluding the practical perspective, by being sceptical as to a person's ability to find the truth using one's own reason. But that was only to be expected.

The difference between MR_{MD} and MR_{MI} is that between an internalism and externalism in epistemology: what justifies our epistemic practices is accessible to us from the first-personal point of view *vs.* what justifies them is thus inaccessible (Zangwill 2012: 351). However, moral ontology and epistemology are mysterious enough to make the externalist justification extremely complicated if not impossible. But again, it is obvious that the justification part is not as important for MR_{MI}, i.e. the part "how we know that what we believe is morally right is actually morally right?" We know what is right more or less in virtue of common sense, so it is not their task. They know moral reality exists and is accessible to us. What needs justification is not the

moral knowledge or its accessibility, but the status of moral knowledge – that it is knowledge, thus, that it is objective.

Further failures of MR_{MI} to harmonise the practical and the theoretical. Meta-ethics for the MR_{MI} is all about explaining how moral knowledge can be objective and how ethics can be a scientific discipline granting moral knowledge. It seems that for $realists_{MI}$ it is just one more sphere that needs to be conjoined in order to complete the scientific worldview. They seem to advocate the view that the only way for morality to be saved from chaos and illicit doubt, is by getting for it the status of science (is the aim of creating a unified science forgotten?).

But the project of scientific ethics can be seen as the effect of aiming at harmonising our various beliefs, i.e. people's theoretical and practical beliefs. Adherents of the MR_{MI} strive towards a unified worldview through one epistemological pathway to knowledge: we acquire knowledge through our senses – empirically. However, there are many philosophers who claim that "the view from nowhere" (Th. Nagel's term) or the "absolute conception" (Williams's term) – the aspiration of a scientific inquiry – in practical realm is impossible. Even if those theorists agree that such an ambition is proper for natural science, it is not appropriate for ethics.

Williams gives several reasons for it. The first is difference of practical and theoretical reason. The second is the differing importance of the person (or the first-person perspective) in scientific and ethical thought, i.e. people are not special in the first, but essential in the second: "The aim of ethical thought, however, is to help us to construct a world that will be our world, one in which we have a social, cultural, and personal life" (Williams 2006: 111). Finally, according to him, it is impossible to commit ourselves to thinking about our ethical life from only the theoretical perspective using the *concepts* available to it.

The problem is that scientific worldview – even if in some way it contained human-independent moral values – does not serve us as a practical

guide. And the supposed harmony is achieved not by reconciling the two viewpoints, but rather by attempts to eliminate one of them by reducing it to the other. But on what grounds is it eliminated? The attempts to reduce moral properties to natural properties have been so far unsuccessful (Vasilionytė 2012, Rosati 1995, etc.) and, as Williams points out, it is still an unsolved problem how much even of the psychological vocabulary could possess that absolute character. Reflection is important in human lives, but so is acting, and acting, as noted by René Descartes and others, cannot be suspended even in face of theoretical doubts, even in situations of uncertainty and limited knowledge. As Williams says, "The only serious enterprise is living, and we have to live after the reflection; moreover (though the distinction of theory and practice encourages us to forget it), we have to live during it as well" (Williams 2006: 117).

Another way to seek the harmony is to acknowledge that there may be two different modes of looking at the world. As for the workings of the non-human world, the theoretical perspective may be preferable, however, privileging the perspective purified from humanity when trying to understand the human world may be unacceptable. One surely cannot posit non-natural properties, because that is what contrasts with the metaphysics of a sensible contemporary theorist, there has to be harmony. But eradicating certain beliefs from the theoretical worldview that one cannot but hold from the practical perspective does not harmonise the person, but puts one into contradiction. As William nicely puts it, the reason does not drive us beyond humanity; humanity requires us to assemble as many resources as we can to help us respect it (Williams 2006: 119).

2.3. Positing the unnecessary ontology

Objectivity with and without objects. But if the already mentioned dichotomy of scepticism/realism$_{MI}$ is false, if morality can be saved without positing the moral objects in virtue of which moral concepts can be objective,

what is it? Can there be, to use Putnam's notions, "objectivity without objects", and "ethics without ontology" that withstands sceptical blows? There are theorists that answer to the positive.

Arguing for objectivity without objects, Putnam gives an analogy with mathematics:

> "Everything about the success of mathematics, and the deep dependence of much contemporary science, including physics, but not only physics, *on* mathematics, supports taking mathematical theorems as objective truths; but nothing supports taking mathematical theorems as descriptions of a special realm of 'abstract entities,' and nothing is gained, in philosophy of mathematics or elsewhere, by so doing" (Putnam 2004: 67).

Therefore, acknowledging the possibility of objectivity without objects enables recognition of statements that are not descriptions, and still within the "range of the notions of truth and falsity" (Putnam 2004: 77). In other words, it is a possibility for a similar construal of objectivity in the moral sphere.

The problem with the concepts like "objective", "real" and the like lies certainly not within them (there is nothing wrong to ask if morality is objective), but rather with the wish to import whole conceptions of objectivity and reality into the domains that are significantly different. The effects of such introduction of the standards that cannot be met is unfair evaluation of the domain: one should either acknowledge it is not up to standards (morality is not real, it is not objective; this is what expressivists maintain) or one can try to save it arguing that it does meet the criteria by slightly modifying it (perhaps even not reflecting that the criteria are unfit *for this exact domain*). These concepts of "objective", "reality", which are loaded with conceptions of objectivity and reality, pose false dichotomies of realism/anti-realism, descriptivism/expressivism, objectivism/subjectivism, and the subject matter falls through the cracks of this Procrustean bed.

Even if Moore, the so-called father of meta-ethics, gave birth to the discipline in anti-reductionist spirit and aiming to bring ethics back into the

circle of proper disciplines and grant it a distinct object of investigation, he was a child of his time and tradition of philosophising, thus, he brought empiricist standards of objectivity and reality with him. As Korsgaard noted, moral realism was often born as a reaction to impeding loss of moral knowledge; and knowledge was defined rather strictly at the beginning of the XXth century analytic philosophy.

One can see that the empiricist criteria for objectivity and reality were way too often taken for granted even by the critics of realism$_{MI}$ and the people who have tried to overcome it: the same criteria have formed realism$_{MI}$, scepticism and expressivism depending on the answer to the question if moral language and moral reality can meet them. Mackie's views can nicely serve as a proof. Mackie held that values were not objective in the sense of being part of the fabric of the world. In his "argument from queerness", he says that "the idea of ontologically objective values explains nothing and offends against parsimony" (Williams 1985: 203).

Williams, on the other hand, does not take those empiricist criteria for granted, so Mackie's views seem to him rather cryptic: why call oneself a sceptic in such a context, why demand morality to be part of the fabric of the world at all? In Williams's view ethics is such an area which is not concerned with knowledge and it is inappropriate to seek it (Williams 1985: 204). However, according to Williams, from the way Mackie uses the term of moral scepticism in his work, one understands that for him scepticism is concerned with knowledge or the lack of it. Both Mackie and Williams would agree that there is no moral knowledge as such, but probably would not agree on the question of whether ethics should meet that requirement in order to be a proper topic of serious investigation.

An ontological or a conceptual/logical matter? As mentioned, it seems that it is a wish to save morality that brings with it a need for moral ontology. Or, alternatively, MR$_{MI}$ seek to have a unified worldview and in order to neatly complete the picture they finally turn to ethics and incorporate it. But it may

very well be that moral questions do not require metaphysical answers. With regard to that, I find Hare's paper "Ontology in Ethics" particularly incisive, where he argues against what we call MR_{MI}.

Hare claims that some ways of posing an issue have advantages over others, because posing it in certain terms would leave it undecidable. And one such wrong way of putting the issue in moral philosophy is putting it in metaphysical terms (the moral realism/anti-realism debate). According to him, ethical realism consists in a view that moral qualities and moral facts exist *in rerum natura*, so if one says that some act is wrong, one is saying that there exist this quality of wrongness and it is in virtue of its existence that the act is wrong. That is, acts have moral properties because of the existence of moral properties.

However, there are different senses of the word "exist" and different kinds or orders of existence. Hare suggests that one can say that "numbers exist in a different sense from cows, or that the existence of numbers is a different kind or order of existence from that of cows" (Hare 1985: 41). And even if one does not admit the need for these different kinds or orders, it still is useful to distinguish between the different senses of the word "exist". These, as Hare puts it, "formal senses"[26] (defined in terms of what we can and cannot rightfully say) of "exist" may be a better way to discuss the issue. Hare says it is probable that even most (of the so-called) anti-realists admit of the existence of wrongness in one of these formal senses, that is, that wrongness exists in *some* sense, but in a different sense from cows. And if those questions that are discussed enable clearer distinctions when put in conceptual or logical terms, perhaps those just are the most perspicuous terms.

In other words, Hare sees no reason for the so-called realist and anti-realist to quarrel. Because if the anti-realist thinks that "even if wrongness exists in *some* sense, it exists in a different sense from cows" (Hare 1985: 42),

[26] One of the senses is such that something exists just in case we can *meaningfully* say that, e.g., something is red, second – if we can *truly* say that, e.g., something is red, and third – if it can be *referred to* (Hare 1985: 41).

and a realist believes that wrongness exists in the same sense, but is a different sort of thing from cows, there is no real disagreement. As Hare rightly observes, provided that some acts are wrong, there is no reason for an anti-realist not to admit that wrongness exists, where this "exists" is used in some weaker sense which fits numbers *as well* as cows.

One should underline that Hare's argument is not fuelled by the wish to dissolve all ontological questions turning them into linguistic questions. It is rather motivated by the wish to disambiguate the main term of the distinction which only obfuscates what is at stake. Indeed, if there was a talk either of different kinds of existence, or – even better – of different senses of "exist", the current distinction or realism/anti-realism would crumble, the purported distinction that excludes some theorists from the party of moral realists would be corrected. It is unfortunate that in the current debates of realism/anti-realism the dominant usage of "exist" is such that does not allow for difference of senses/orders.

For example, many of the constructivist theorists, as well as, for example, Blackburn, identify themselves as anti-realists. Blackburn calls his theoretical position "quasi-realism". But they do so only because the realist position is supposed to mean that "when we moralize we respond to, and describe, an independent aspect of reality" (Blackburn 1985: 11). However, I believe it is misleading because this particular strand of anti-realism – in arithmetic as well as in ethics – holds that the practices at hand are "as solid and certain" as they can be without being explained by reference to some independent (mathematical and, respectively, ethical) reality (ibid.). As Hare puts it, an anti-realist "can easily agree that they [moral properties and facts] exist in some other senses, or even exist *in rerum natura*, if that term is taken more liberally" (Hare 1985: 48).

However, as Hare notes, a philosopher who affirms that moral properties or facts exist (in some sense), "has done absolutely nothing to solve the main problem, namely how we determine that they exist in a particular case … how to determine … that an act is wrong" (Hare 1985: 49). And that, as we have

seen from Street's argument, is something MR_{MI} has especially hard time with. Then, if we should not worry about whether morality exists in some sense (because it does), Hare thinks that the realist would be unwise to claim that moral qualities exist in any stronger "material" sense and that he does not need to. But whether he is right about the latter point or not, we should rather concentrate on what Hare calls epistemological or logical issue: "how to give an account of moral thinking which allows to arrive in a rational way at conclusions which are practical and prescriptive" (Hare 1985: 49).

Why not to remain with the one and only meaning or order of existing? The current situation when the ontological distinction is based on one randomly and silently selected, or, more probably, historically set meaning of "exist", while it is an ambiguous term, makes part of the moral arguments of the supposed adversaries murky or pointless. Besides, to consider the case of numbers, we would either try to make numbers meet the criteria of reality fit for cows and get stuck in futile philosophical equilibristics mutilating or transforming numbers into something that they are not or what is an incomplete picture of them, or we will have to acknowledge that numbers plainly do not exist (do not meet the criteria). But why not rather acknowledge that they exist is some sense different than cows without positing any suspicious (non-naturalist) ontology and give a good explanation of their mode of existence (probably through a plausible epistemological-conceptual story).

The narrowness of thought which does not acknowledge differences in order or sense is apparent not only in case of "exist" or "reason" or "rationality", but also concerning "knowledge". Hare claims that knowledge can consist of beliefs different than factual. Moral knowledge is not about facts, it is obtained while determining the questions rationally (the descriptive part of moral sentences is never enough). "Beliefs" can also refer to such mental states which are different from factual beliefs. Thus, acknowledging that one of those connected terms is ambiguous brings with it a need to acknowledge the ambiguity of the rest of them.

Besides, even if it is apparent that moral properties or entities are different from cows, it does not mean we have to accept the analogy of morality with mathematics. However, it seems that we face a choice of one or another type of analogies, as MR_{MI} also makes use of analogies (even extensively) when striving to establish their point. They rather use analogies of moral and non-moral (reducible or irreducible) natural properties (such as redness), natural kinds (e.g. water) and so on. Their opponents, respectively, try to discredit these analogies and propose mathematical ones instead, claiming that the latter ones approach the character of moral properties closer than the former ones. So we will take up the question of analogies (if they succeed) separately and shortly.

Does moral language presuppose a particular conception of objectivity? What, in my opinion, confirms the suspicion that in ethics the central questions are not those of metaphysics is the analysis of Mackie's error theory. What it shows, is that the necessity to present a certain moral metaphysics is introduced by semantic views. If the latter are such that truth conditions of moral language require existence of moral objects, it becomes inevitable to explain how this moral reality is possible. However, having a different semantic theory may absolve from the need of moral metaphysics in the sense of a theory of a mind-independent reality.

When reading Mackie, I always sensed that there was something more than those two explicitly stated claims that allowed for his conclusion about the error involved in our moral linguistic practices. Namely, that there was a hidden supposition that enabled this exact conclusion and that without it, the conclusion would be toothless against some variants of moral language usage, such as, for example, a constructivist variety. Surely, I was not the only one to notice it and it was Blackburn who managed to show nicely what this supposition was (or what explicit refinement/restriction was missing from the actual Mackie's argument). Mackie supposed that objectivity was definitive of ordinary moral terms, with which one could agree: people indeed speak as if

morality was objective. However, what one cannot agree with is a certain conception of objectivity that he ingrains into the common moral language[27], or, in other words, a certain mode of existence of those moral values.

Blackburn claims this supposition (that exactly this defined conception of objectivity is presupposed by the moral language) is false and gives an example. Mackie himself, though recognised an error, went on to use the same "infected" language which served the practical needs of people to moralize, however, without the bad ontology. Blackburn humorously calls it "shmoralizing" to mark the difference between the two practices. He considers for what possible reasons the moralizing and shmoralizing could have seemed identical, but the only plausible answer was that they were indeed identical. That leads to the conclusion that if from the practice you are unable to tell which moral ontology people adhere to, the moral practice does not imply any, so it can be "clipped on to either metaphysics" (Blackburn 1985: 4).

"Either metaphysics" is not just whatever metaphysics, of course, but it means that moral semantics does not presuppose one and only metaphysical view (as well as one and only conception of objectivity). And it is exactly the opposite that Mackie relies on in his argument. Then, Mackie moralizes "just as ordinary people do, but with a developed and different theory about what it is that they are doing" (Blackburn 1985: 3). And the error theory, according to Blackburn, shrinks to a claim that most ordinary moralists have a bad theory about the moralizing and about the objectivity of values that they believe to be involved in the practice. Thus, the idea of objectivity *is* embodied in the moral language, but its exegesis is a matter of a further debate.

Street is of a similar opinion. She sees Mackie's argument, already supplied with a certain conception of objectivity, as successful, but limited in its scope: "offering reasons to think there are no objective values in the realist's sense, Mackie says nothing at all to address the second debate, thereby

[27] Another indicator that moral ontology may be unnecessary for a proper functioning of morality, as well as for grounding its claims, is that common-sense practices while having the cognitivist and internalist suppositions do not have realist$_{MI}$ suppositions.

leaving open the possibility of objective values in the Kantian metaethical constructivist's sense" (Street 2010: 379). According to her, then, a person has another choice than that of defending realism$_{MI}$ with its metaphysical and epistemological difficulties, and that can be a Kantian meta-ethical constructivism (ibid.).

There is, of course, a different opinion. Williams, for example, does not charge Mackie either with importing certain suppositions or with a wrong conception of objectivity of morality, but accepts Mackie's supposition that we experience objectivity of moral demands as a robust objectivity as understood by moral realists$_{MI}$ (and, thus, Williams accepts an error-theoretic conclusion). He rather uses Mackie's argument against the target that he finds more significant.

In moral philosophy it is well known that Williams's relation to the moral philosophy of Kant is complicated. On the one hand, he holds Kant to be right about many things. And tackling the question at hand Williams points to alternative theories of objective morality, such as Kant's, stressing that they deploy "an intelligible and adequate sense of objectivity". The objectivity is not obtained through the relation of the statements and the world, but, in Kant's case, "through the relation between accepting those statements, and practical reason" (Williams 1985: 206). Kant is not a realist$_{MI}$, but offers an "objective grounding" for morality ("objectivism as a kind of realism", Williams 1985: 208) which seems to be a more suitable alternative to objectivity$_{MI}$.

However, on the other hand, Williams is a harsh critic of Kant. And acknowledging Mackie's supposition that language users perceive moral objectivity as objectivity$_{MI}$ lets Williams conclude that "[t]here would be a certain misrepresentation, then, in that experience of objectivity even if there were genuine objectivity in the form of an objective grounding" (Williams 1985: 207). Thus Williams extends the grip of Mackie's argument onto the realist$_{MD}$ objectivist moral theories, because, as Blackburn has noticed, without accepting the aforementioned supposition Mackie's argument would remain

toothless: objectivity of moral language could be accounted for in some other way – some such way than he did not presuppose. So Williams holds that any objectivism about moral language is implausible and Mackie was right to reject it even on wrong grounds.

However it may be, one should be aware of the possibility to have different conceptions of objectivity, and that Mackie's argument at least poses a challenge to the MR_{MI} construal of objectivity. Mackie, criticising such moral realism correctly targeted their conception of objectivity and its ontological and epistemological implications. It is another question, though, if exactly this conception is engrained into the common moral language. Blackburn shows this supposition to be doubtful and I find Blackburn's example compelling.

2.4. Failed analogies

Disanalogy with colour terms and terms for natural kinds: semantic differences. In moral philosophy two tools that are used rather pervasively, but the result of which cannot be taken for granted, is metaphor and analogy. Both are formed in accordance with, respectively, the hidden or apparent similarities of some two things and both help to inspire the feeling of understanding the subject matter better: given one understands how one thing works/is, presuming the other thing is relevantly similar, one understands how the other thing works/is. Sometimes they are almost inevitable because we know more about the thing that we draw analogy with than about the subject matter itself and we want to add more certainty to what we say about the subject matter. Proposing an analogy is like *sketching* a positive proposal, which we cannot fill in with details, in other words, it is like gesturing in the direction of the explication, suggesting that it should go along these lines.

However, similarities can be inessential, or the differences can be relevant enough to dismiss the analogy when talking about some aspects of the subject matter. Similarly is with metaphors: they may be impressive, but when unwrapped, they prove to be a rhetorical means rather than a philosophical

vehicle. That is why I find it useful to tackle the main ones to this debate. One such analogy of moral properties is with natural, so-called "secondary", properties, such as colour properties. Another significant analogy prevalent in the context of semantic reference is that of water/H_2O and good/subvenient natural property (this is just a placeholder for a specific natural or at least naturalistically describable property). The opponents of the colour analogy, though, try to replace it with a number analogy (on both semantic and more substantial levels).

To turn to the question of which analogy – that with natural properties or that with mathematics – fits moral properties better, there is plenty of evidence that natural non-moral properties have just too many differences from the moral properties for the analogies to do the required work in moral philosophy. A usual analogy is that with colours, as both colours and moral properties supervene on naturalistically definable properties.

For example, Hare[28] explores the analogy of moral properties and facts with "ordinary" properties ("secondary" properties as termed by Locke) and facts, such as a property of redness and a fact that a thing is red. It seems primarily that the analogy is apt: both are "the joint product of properties of the object or act (which are themselves not identical with the quality of wrongness or of redness) and of reactions in the perceiving or thinking subject" (Hare 1985: 46). However, there is an important difference that Hare brings out. If somebody says that a thing is red, whereas those conversant with language and viewing that thing under normal conditions do not call it red, that person is either making a mistake or is colour-blind. However, if somebody calls an act wrong whereas other people, equally well informed and conversant with language, claim to the contrary, it is not necessarily true that that somebody is either morally blind or does not know English. It may very well be that the

[28] Even if Zangwill is a moral realist$_{MI}$, Jackson is a moral realist of a functionalist variety, Blackburn is a quasi-realist and Hare is a prescriptivist, I take it that they have important lessons to offer. Here I will use their negative part of the theory, that is, their criticisms of some strands or tendencies within realism. However, I find their positive suggestions flawed in one sense or another, and at that point our paths part.

majority is in error, say, about eating meat or abortion (as it was about slavery, women's rights etc.).

Another analogy is based on the functioning of such a word like "water" in our language. The Twin Earth thought experiment by Putnam that I will present concisely appeals to intuition of speakers and it goes like this. The word "water" for us, Earthlings, denotes a certain natural kind, the nature, or essence, of which is determined by its microstructural properties (H_2O). In other words, for us "water" rigidly designates H_2O. On the Twin Earth, though, there exists a liquid which shares the same manifest properties of our water (drinkable, colourless etc.), is referred to as "water" by the Twin Earthlings, but which has a different microstructural essence (XYZ). So in case "water" for a human being rigidly designates other microstructural properties than it does for an inhabitant of the Twin Earth, the intuitions of the competent language users are such that the two beings have no reason to disagree about what "water" means: the extension based meanings of the two homographs "water" are different.

Therefore, if "good" functioned in the same way as "water" did, the Moral Twin Earth scenario would yield the same results. However, some of the contemporary philosophers argue that moral terms do not function the way "water" does. As another of Hare's examples (1991/1952: 148-149), a famous cannibal island case[29], shows, as well as an extended and more explicit Horgan and Timmons's Moral Twin Earth example (2006), the *moral* argument of the missionary and the cannibals or that of the Earthlings and the Twin Earthlings (to the contrary of the argument about water) is not silly, but genuine; the ground for an argument exists, it is there.

The *moral* disagreement cannot be dissolved by the different sides acknowledging that they give different natural properties the same name. Maybe the Earthlings and Twin Earthlings would need to use some kind of indexing when they talk with each other and need to point to the different

[29] This example is also made use of by Michael Smith (1994: 33-35). I will tackle only Hare's and Horgan and Timmons's views here.

substances called “water” – in so far as that has implications for practice, for example. However, in so far as they talk to each other using the words “good” or “right” for practical purposes, indexing does no good. They cannot happily acknowledge that “good” or “right” are just names, like, for example, “Ann”, which point out to different people; at least one side should be wrong calling “good” or “right” that which it does.

But if both sides are thought to be right calling different things “good” and “right”, then, as Hare rightly notices, moral realism$_{MI}$ along with its analogy condemns morality to relativism. I.e. if “right” and “good” are terms of recommendation in which one can be objectively right or wrong (if these are “the manifest properties” of goodness and rightness), then a missionary’s or an Earthling’s “good”, designating the meek and gentle, and a cannibal’s or a Twin Earthlings “good”, designating, e.g. those who collect the most scalps or those who are green and aggressive, would just be homographs. Thus, there could be no real disagreement about what “good” means: the extension based meanings of the homographs would just be different; there would be no common subject matter of the conversation.

The message of these counterarguments to the analogy of moral terms with terms for colour and for natural kinds is that semantic theories that apply to the latter do not apply to the former: moral language functions differently.

Still another semantic difference is spotted by Blackburn: moral practices vary with the forms of life of a society in a radically different way in comparison to the perceptions of secondary qualities. A predicate expressing a certain moral property can get a radically different extension, for example.

Also, Blackburn joins Wiggins in noticing that evaluative predicates are typically attributive, but not those of secondary properties: a man may be a good burglar, but a bad batsman, however, a red tomato is a red fruit and a red object bought at the grocer’s (Blackburn 1985: 15).

Realism$_{MI}$ is bound to negate the inaptness of the analogies. Moral realists$_{MI}$ usually do not acknowledge there is a difference between goodness and redness or being water if they see all terms as *denoting natural features*.

That is, it is not that "good" or "right" recommend or require something in virtue of certain natural features, instead, they denote those natural features in one way or another. It is for some of their critics that "good" or "right" express recommendation in virtue of something, i.e. they view an argument between different parties as an argument about what is worth recommending (or is justified to require for somebody's sake). For realists$_{MI}$ moral "arguments" do not really deserve the name of an "argument" as there is nothing to argue about: there should be a way to determine who is misusing the word "good" or "right" (not by weighing rationally the arguments of the differing sides) or who has a flawed moral perception. From their point of view the causes of being mistaken are flaws in linguistic competence or psychophysiological make-up (or of knowledge about the natural world), not flaws in reasoning.

It is a good place to notice an important thing. From the point of view of the MR$_{MI}$ moral constructivists and the like are using concepts such as "knowledge", "exists", "natural", "true" and the like in a weird way, because the dominant usage of these is that of the "theoretical" philosophy. However, from the constructivist point of view *and* from a common-sense point of view, it is MR$_{MI}$ that is using terms such as "reason", "justification", "argument" and the like attributing them unusual, weird meanings, because these are the terms that make sense in the field of "practical" philosophy.

Here I also want to stress that whereas usually Hare's and Horgan and Timmons's arguments are thought to be harmful to moral realism and thus to support expressivism, I claim that this in not necessarily true once we drop the overly simplistic dichotomy of moral realism/expressivism. Nor one needs to accept the detailed positive proposition of Hare in order to accept his criticisms advanced to MR$_{MI}$ or to approve of his general guidelines for the correct moral theory.

Epistemology related semantic differences. Whatever the particular ontology, MR$_{MI}$ advocates a perceptual approach to morality: mind independent reality is out there and we perceive it in certain ways. However,

the perceptual account of epistemic paths to moral reality is implausible for several different reasons. Blackburn gives several reasons against the analogy of moral properties (and values or obligations more generally) and colours in this respect. Some of them come from the phenomenology of morality that is partly acknowledged by moral realists$_{MI}$ themselves, other reasons are that it leads them to bad philosophy.

First, moral properties and secondary properties supervene[30] upon primary properties in a different manner. Whereas, according to Blackburn, it is a scientific fact that secondary properties supervene on the primary ones, it is not in the moral case. In other words, we know only *a posteriori* that secondary properties supervene on primary properties, but it is *a priori* that moral properties supervene on the natural ones. The metaphysics is not important here, the difference is epistemological, and it is apparent that senses do not serve for getting knowledge *a priori*. A similar difference is brought out by Jackson: "[t]he supervenience of the ethical on the descriptive is, by contrast, *prior* to metaphysics" (Jackson 1998: 128).

One could add to this epistemic disanalogy a metaphysical aspect brought out by Zangwill. He claims that the metaphysical asymmetry consists in there being no essential relation between the moral and natural properties. Zangwill draws attention to Kit Fine's distinction between essence and modality. So for example, water is necessarily H_2O because it is essentially H_2O, but pain is necessarily bad not because it is essentially bad (pain, according to Zangwill, is sufficient for badness, but it is not in its nature to be bad) (Zangwill 2005: 127). Therefore, it is not essential to natural properties to generate moral properties, so supervenience is not a relation of essence or identity in moral philosophy.

By the way, a constructivist, or a moral realist$_{MD}$, can neatly account for why "it is not essential to natural properties to generate moral properties":

[30] The idea of supervenience here consists mainly of the idea of covariance, and is a further question in virtue of what the covariance relation holds – nothing is presupposed in this respect.

because it is not in the essence of natural properties, but in "essence" of the norms to generate the moral properties. Moral properties are instantiated when our normatively restricted look meets the natural world. MR_{MD} can explain the supervenience relation as well: it is the same norm that enables the same moral properties to be instantiated when faced with the same natural properties.

Given what I said, one can make an even bolder claim: in view of the fact that both MR_{MI} and MR_{MD} that accept this (rather than the one fitting for the natural kinds) conception of supervenience, an explanation of how moral values come about provided by MR_{MD} is superior to an explanation by MR_{MI}. For example, Zangwill answers the question of why pain is necessarily bad with "the conjunction of two facts: firstly, the fact that it is an essential property of all moral properties that when instantiated there are some natural properties that suffice for their instantiation; and secondly, the plain fact that pain is bad" (Zangwill 2005: 127). The explanation does account for the supervenience relation, however, a further question arises of why pain is bad (especially if it is not, according to same Zangwill, in its essence to be bad). Zangwill understands that and rushes to answer with "it just is, and … explanations must come to an end somewhere" (ibid.). The MR_{MD} does not need to cut the explanation short or to rely on some "right" mode of perception that enables people to see what is right. Supervenience for MR_{MD} does not cover any inexplicable mechanism – it is apparent how it works. The MR_{MI} cannot answer this in principle: the ways of the nature do not need to make sense, the "why" question is inappropriate in cases like this or when we ask why a molecule of water consists of the atoms that it does.

Second, Blackburn notices that the receptive mechanisms which are responsible for our acquaintance with secondary properties are well-known objects of scientific study, whereas the studies of what defects lead to moral blindness are not at all similar (Blackburn 1985: 14). No sensory, receptive or causal mechanism is the topic of such studies. Moral blindness, according to Blackburn, is not immediately accessible to the person oneself – while it is to the one whose secondary property detecting mechanism fails. I suppose that he

talks about such cases as, for example, one's seeing world in colours in some situations and then becoming colour blind in other situations, or failing to hear anything when that is to be expected because of one's previous experiences. Whereas one's failures in moral judgements are not like that: we do not know as clearly if we "perceive" the moral world or fail to in various situations.

Third, the mind-dependence of secondary qualities and moral qualities is different. To use Blackburn's example, if we changed so that what appeared to us blue came to appear red, the world would cease to contain blue things, but if everyone came to think that it was permissible to maltreat animals, it would not stop being bad – it would only mean that everybody has deteriorated[31] (Blackburn 1987: 14).

If all this is so, then the perceptual account in relation to morality comes to doubt. The most disconcerting is the problem that the apparent epistemic pathway to moral properties or truths in not that of perception. To start with, we do perceive something as red, but the phenomenology of morality tells us that goodness is not something we perceive, but is rather subject to reasoned choices (Hare 1985: 47).

Indeed, as Blackburn notes, the ethical typically concerns imagined or described situations, not perceived ones, because of its action-guiding function. We reach verdicts in light of general standards which are also not perceptually formed or maintained (Blackburn 1987: 365). The possibility of generalisation, according to him, is also rather implausible: "How could I be sure of the

[31] This point may be understood as targeting the idea of mind-dependence of the moral properties. However, I see it as damaging not the idea itself, but the *passive* or the *receptive* character of this dependency. I.e., moral properties are there to be perceived regardless of whether we do, but the explanations of why and how they are – regardless of our perceptual abilities – will differ as to their plausibility. MR_{MD} will be, once again, able to give a non-mysterious explanation: they are there potentially, or in relation to an idealised rational person's mind. The story of *everyone* loosing this "perception" and something being bad is not even conceivable. But if we can say that despite our perceptual abilities to sense it, we know that badness would be there, *how* do we know, what epistemic path led us to this knowledge? The answer may be simple enough: reason. But the perceptual story of a rational way of gaining of ethical knowledge, that is, the story about intuition, fares no better: it blocks explanation and our understanding of the moral domain; thus, moral ontology and epistemology are again just a matter of faith.

generalization to examples I did not see (I could not do that with colour, for instance …)?" (Blackburn 1987: 365).

To go further, as Smith points out (in Smith 1994: 23-24), if causal contact with the natural state of affairs is not necessary for determining the moral value co-instantiated with it, then moral knowledge is not explained by a causal perceptual story. Thus, moral knowledge is not a type of causal knowledge. The analogy with perception of the natural world does not hold. Even if with slight difference, these criticisms apply to both the naturalist and non-naturalist versions of MR_{MI}.

So it seems that terminology of a passive perception and *a posteriori* knowledge is not acceptable in the moral domain, but is perfectly fine when talking of colours and similar non-moral properties. If, as was noted before, the epistemological/logical question is indeed central to the moral debates, this should give MR_{MI} a pause. It is true that in many cases moral realists appeal to the moral perception, or sight, the correct way of seeing the moral values. However, if the moral phenomenology is essentially different from non-moral phenomenology and moral language functions differently than non-moral language, the burden of proof that these differences are irrelevant is on the shoulders of MR_{MI}. And the naturalist variety of the MR_{MI} is in no better position at explaining how we come by the moral knowledge than the non-naturalist MR_{MI}.

Disanalogies in practical implications. Another difference, according to Blackburn, is the necessary practical implications or a lack thereof. Blackburn notes that it is up to a subject if one cares about any particular secondary property, but the practical nature of morality is intrinsic to it. Otherwise there could be a theoretical space open for a culture which perceived the moral properties, but paid no attention to them (Blackburn 1985: 15). However, he claims that this space is closed. Even if one could challenge this latter claim, one can clearly see the difference between the secondary qualities and moral properties in this respect.

Another stark difference is pointed out in Williams's work where he discusses error-theory. Discovery of the fact that secondary qualities are subjective have little or no effect on our everyday practice. The unreflective practice and the theoretical understanding are harmoniously related (Williams 1985: 211). Williams even thinks one can make a good case against the view that there is an error involved in everyday belief. However, coming to know the truth that ethical qualities are dependent on us whereas it feels that they do not would not be harmonious. There can be a story told of why we should stick to or would rather believe that ethical qualities are objective and internalise them, but then one should accept this pragmatic reason (for the sake of the smooth working of the society, for example). Williams reports that Mackie himself wrote that "if subjectivism were not just true but known to be true, those processes would be consciously conducted in some different way" (Williams 1985: 212).

One way of taking this point is to suspect that whereas in the colour case our practices are rooted in our perceptual make-up that is out of our wilful control, in the case of moral practices we are dealing with different mechanisms. The very *possibility* of practice change (that in face of the truth that morality is subjective the practices *could* change) shows that it is not totally out of our control.

So one can already sum up the considerations against the analogy with natural non-moral properties. Blackburn claims that these considerations have a cumulative effect against the advocates of an analogy with the secondary qualities, as well, I could add, as that with other natural properties. He, I think, rightly claims that it provides no real explanation or theory, but only a misleading sense of security that there is such a theory to be had, whereas it is barely gesturing at a lame analogy (Blackburn 1985: 16-17).

As for the analogy of morality with mathematics, I believe, it is only needed for the MR_{MD} to open up the space for a different kind of conception of objectivity – to obtain which no ontology may be needed. The theorists that advocate $realism_{MD}$ do not rely on this analogy more than needed, because they

can propose a clear and explicit epistemological-conceptual account of how moral judgements can have truth values and how we know in particular instances that they are true or false. That is a difference: the realists$_{MD}$ can use the analogy with mathematics for a negative work and then shed it when the positive part of their endeavour starts, whereas the realists$_{MI}$ rely on the analogy with natural kinds and secondary properties all the way (but it turns out that the analogies are unable to do what they intended).

2.5. Implausible semantic theory of normative terms

Normativity revisited. To remind, Zangwill suggested that MR$_{MI}$ justifies the authority of morality metaphysically, not from the first-person point of view. So in order to be just to the MR$_{MI}$ theories, beside posing a normative question in terms that do not allow the MR$_{MI}$ to answer it (by Zangwill's lights), I evaluated the plausibility of the MR$_{MI}$'s moral metaphysics and epistemology in their own right. The evaluation revealed that the moral ontology hangs on faith, whereas moral epistemology is doubtful at best.

The authoritativeness of morality, then, is a serious problem for the MR$_{MI}$ as moral ontology cannot justify the normativity of morality – it is a matter of faith itself. However, there is one more test, and this time it is a test for any *descriptivist* theory. This test, the Open Question Argument (OQA), once used to argue for the non-natural character of the moral properties and later – to argue for the expressivist character of moral judgements, now can be understood as a test of normativity: can any descriptivist theory of moral terms retain their normative character. So we revisit the question of normativity, but from a different angle, that is, putting the varieties of the moral realism to a more or less neutrally formulated test.

The OQ(A) as a test of normativity for descriptivist accounts. Moore's Open Question Argument (OQA) was to expose the Naturalistic

fallacy involved in identifying the good with any particular natural property. However, nowadays the scope of the OQA is understood not necessarily as that of compromising naturalism, but descriptivism more generally or reductionist naturalism more particularly. This is well illustrated by one of the interpretations of the OQA[32].

Some theorists think that Moore's Open Question is best understood as a test of normativity. This test is particularly hard for descriptivist accounts: how can a description be normative? It is tricky to talk about "descriptivist" accounts, though, because, as we have seen, some of the cognitivist accounts are descriptivist in a different sense than the dominant one. However, here I will use "descriptive" in such a way as to include both senses of "descriptivist" and, respectively, "descriptive", unless noted otherwise[33].

This interpretation of the OQ relies on moral phenomenology: what an adequate theory of moral terms needs to capture is twofold. On the one hand, we have to acknowledge that moral phenomenology encompasses a relation between moral properties and natural properties. On the other hand, moral terms are commending, or, as it is put today, normative. Therefore, descriptivist theories that reduce goodness to some or another natural property (probably *in virtue of* such a reduction) usually leave the normativity out, and that is exactly what the OQ test is used to indicate. If "good" simply denotes a natural property, it cannot be a recommendation by default: "when the concept of the good is applied to a natural object, such as pleasure, we can still always ask whether we should really choose or pursue it" (Korsgaard 1996: 43).

But even if the OQ purported to show that no naturalistic description or analysis of the good could capture the essence of moral properties *completely*, that there was always something *essential* left out, that the essence of moral properties could not be captured by their actual descriptive definitions, it would

[32] There are more readings of the OQA; for more see (Vasilionytė 2012).

[33] If one uses "moral realism" to refer to a broader position which can be realised differently, then one should do the same with all related terms, such as "objectivity", "descriptive", "knowledge" and the like. One can use indexing, when needed, to refer to the more particular conceptions of the said concepts.

not mean that this OQ is in principle hostile to naturalism about moral terms. For example, Rosati thinks that the OQ is not a tool for *undermining* naturalism (or descriptivism) as such, but that it is an important tool *for* a naturalist herself, "a device for unearthing features of our ethical concepts, or better, of the properties our ethical terms express" (Rosati 2003: 501).

I have to note that Rosati's discussion is primarily that of non-moral goodness, but I believe that the same applies to the extension of discussion on the moral goodness (as moral goodness is still a case of goodness, even if with its own particularities). She reminds that certain "new naturalists", such as Rawls, Brandt, Railton, and Lewis, do appreciate that, despite its flaws, Moore's OQA exposes a problem that any naturalist must address. While agreeing with the non-cognitivists that "earlier forms of definitional naturalism failed effectively to capture the expressive and recommending functions of evaluative terms", they deny this to indicate that evaluative property terms are not purely or primarily descriptive (Rosati 1995: 46). They believe, according to Rosati, it to be possible to "construct a descriptive meaning for 'good' that secures its recommending and expressive functions simply in virtue of the proposed descriptive content" (ibid.). That which earlier forms of naturalism failed to capture (and in virtue of which judgements of goodness are recommending) is normativity. The OQA thus understood, applied to the earlier naturalist accounts, enables to see why the Moorean question remained open, what was that important element that was left out of those definitions and to include it into a new definition[34].

[34] "The new naturalists have identified three specific questions, one or more of which were left open by past definitions of 'good': Does what is said to be good carry motivational force?; Does what is said to be good for a person reflect what that person most values?; Does what is said to be good for a person meet conditions of justification? They have attempted to construct an account of 'good for a person' that closes each question in turn, thereby closing the question whether something that satisfies the account is good for a person. They have not, of course, closed all questions about our multifaceted notion 'good' by closing these questions. But they have, if the new naturalists are right, shown how the narrower notion 'good for a person' can be at once fundamentally descriptive and normative. The worry that

According to Rosati, the OQ owes its unceasing vitality to the very structure of human beings. She says that what explains the force of the OQA is the way we are, the human nature: "The question, then, is not simply what the meaning of good would have to be like (e.g. expressive rather than descriptive) in order for the open question argument to have force but what we would have to be like in order for the argument to have force" (Rosati 2003: 505). She thinks that it is this question that should guide the naturalists. However, as we know, usually the naturalists assign goodness to a natural world narrowly conceived, i.e. to the human (mind) independent reality. So Rosati's claim the new naturalist accounts to have failed as well the old ones comes as no surprise. She believes that they do not manage to include into their definitions what she considers to be an essential element of normativity – an ideal of the person, or, as she later (in 2003) puts it, they (definitions) "do not bear the proper relation to agency" (Rosati 2003: 521).

Rosati notes that such naturalist accounts alienate the goodness from people, as if there was an unbridgeable gap between what people care about, what they want and seek, and what is good (here – even not to say "morally good"). I would say, as if goodness existed in the world the way wateriness did and the recommending character to the "good" would be added accidentally on some occasions – as it would to "watery" in case a human being was practically interested in it. It surely is plausible that a person may not know about one's own good or be mistaken about it or be disinterested in it under some particular circumstances, i.e. it is possible for the motivation to seek the one's own good to be absent. But it seems weird that acknowledging something to be good, especially for one's own self, under normal circumstances can exclude recommendation. Sentences like "pleasure is good" or "vitamins are good for you" may differ in contents and the addressee's reactions (such as willingness to adhere to it based on the truthfulness of such sentences or in motivational levels), but both share the same message: those

---

underlies the open question argument, as the new naturalists interpret it, has thus it seems been met" (Rosati 1995: 52).

are recommendations. Rosati points out the difference between "Skiing is good for Jeff" and "Jeff desires to desire to ski" – the former automatically recommends attitude and action, whereas the latter does not function as a guide (Rosati 2003: 503). (And this difference is an indication that the latter analysis of "good" loses one element essential to the meaning of "good".)

So in case of naturalist reductionist ethical theories normativity seems to be lost between the impersonal identification of goodness with a certain natural property and between the first-person question about goodness of a certain thing in the natural world. Analysing the success of the OQ, one may have a feeling that a person is not asking for scientific tools for identifying goodness in the natural world, but is asking to justify why a certain natural thing with certain natural properties is worth pursuing. In that case a problem with reductive naturalist's ethical theory is that it cannot explain and justify at the same time. It seems that once again MR_{MI} reduces all human enterprises to quests for (theoretical) knowledge, whereas there is reason to think people sometimes need a reliable guidance rather than a solid knowledge of what the natural non-human world is like. In this context Darwall, Gibbard and Railton bring out the practical dimension of goodness as well. They suggest that goodness has a conceptual link to action-guidingness, and it is the dismissal of this link that explains the persistence of the OQA in ethics (Darwall, Gibbard and Railton 1992: 118).

Rosati is rather pessimistic (though not categorically) about the possibility for the naturalists to remedy their definitions "without abandoning their reductive program" in general (even if her criticism was directed to a particular kind of ethical naturalism). Certainly, this version of the OQA does not prove outright that no form of semantic naturalism can capture normativity: an additional argument would be needed to establish that conclusion. It only spreads pessimism about whether it can, based on the failures of the naturalist theories that have been proposed so far and on an observation that it seems to be possible to ask the Moorean question about whatever naturalistic identification of evaluative property.

Rosati's scepticism with regard to ethical reductionism, I believe, is very much in place. One could have wanted to object to my giving example with wateriness saying that we usually talk about pleasurableness and other supervening properties, and not the subvening natural properties. However, that would just be pushing the same problems one step further, because on a reductionist picture same pleasurableness or other supervening property terms would not (necessarily) be terms of recommendation.

From what we have seen so far, even if the OQ poses difficulty for naturalism or descriptivism about evaluative terms, it is rather their stricter versions that have not managed to cope with the task of fitting goodness with human agency. As Rosati notes, the guiding impulse behind naturalism is to develop an account of value that would be continuous with the natural and social sciences, but what equips us for normative life and language, frees us "from the dictates of nature, even our own nature. Still, this feature of us is itself an aspect of our nature – at least as idealized" (Rosati 2003: 524). Thus, "nature" can be conceived so as to include human reality which may be hard to reduce to non-human reality and such version of naturalism may fare better.

The same applies to "descriptivism": if one thinks that evaluative terms, despite their necessary relation to natural properties, do not denote those properties and cannot be reduced to terms for natural properties or natural kinds, then one has a form of descriptivism that may fare better. Therefore, I claim that the OQA poses an especially hard challenge for *reductive* ethical naturalist accounts. One apparent route for a reductionist is, of course, to deny that normativity is an essential feature of moral terms[35], but many of their adversaries are unhappy with such a strategy and keep on pushing them to answer the normative question.

[35] For an interesting account see Strandberg's "A Dual Aspect Account of Moral Language" (2012a) where he devises Paul Grice's notion of generalized conversational implicature to explain the "meaning-like" relation of moral language and practical attitudes, i.e., to show that the practical aspect of moral language is due to the context of the moral utterances rather than being part of their meaning. Though, surely, it is just one more way of explaining normativity *away*.

Rosati's "new [semantic] naturalists" seem to possibly include (value/moral) realists$_{MI}$ as well as (value/moral) realists$_{MD}$ (e.g. Rawls is usually conceived as a constructivist). The formal definitions of goodness given by the "new naturalists" can be supposed to enable an *a posteriori* identification of a natural property (or properties) with the property of goodness. Reductionists expect to be able on that basis to form a reforming naturalistic definition of goodness. Non-reductionists, though, can treat the formal definition as *the* definition or *the* analysis rejecting the possibility to define goodness in natural property terms. The reasons for their reluctance to believe in the possibility of reduction may differ.

A non-reductive naturalist about evaluative properties may think this, for example, because of the multiple realisability of evaluative properties (as, say, it is unlikely that evaluative/moral properties are necessitated by one and only natural property under different circumstances). A naturalist in the wider sense of the term, that is, a constructivist about evaluative properties, may think so because evaluative/moral properties are not necessitated by natural properties: the formal definition of "good" "picks out" different natural properties under different circumstances. Therefore, it seems that naturalism about evaluative properties, or descriptivism, stands a chance of answering the so-called OQ, but only its non-reductive varieties.

Despite the attempts of the MR$_{MI}$ to propose a theory of morality that would preserve its authoritativeness, it seems that it fails. However, we have seen that the MR$_{MD}$ fares much better. But it is time to see in more detail how the MR$_{MD}$ copes with the task. I have given an outline of what their conception of objectivity is, how they understand the task of moral philosophy and what the truth conditions of the moral judgements are. What remains to be explored in detail, is the other characteristics of the moral judgement – its practicality. The authoritativeness of moral judgements, we have seen, consists in their being normative not only by being objective in some sense, but also in their being necessarily action-guiding. And it is the latter aspect that we have not touched yet. So the task of the next part of the text is to explore how plausible

is the internalist premise as embodied in the MR_{MD} theory. That is, if the constructivists manage to give a plausible theory which unites the cognitivist and internalist premises concentrating on the latter.

Part II

Accommodating common-sense morality: action-guidingness of moral judgements

1. The conception and restrictions of motivational internalism

Motivational internalism. The idea that moral judgement has a practical upshot, which is referred to as 'internalism', is rightly called by Smith a vague label. Motivational internalism (MI)[36] in its most general form is defending a claim that there is a necessary relation between a moral judgement and motivation to act accordingly[37]. For example, in a paper by Björklund et al.

[36] Talk about "motivational internalism" should not be confused with talk about "internal reasons" or "motivating reasons" or "reason internalism"; these latter matters, in so far as they are connected to a conception of motivational internalism, will be discussed in this text. However, one should not assume, for example, that the motivational internalism concerns *motivating* reasons just because of the same predicate "motivational" (we will see how they are understood further on); the primary object of motivational internalism is *moral judgement*.

[37] When discussing motivational internalism, some authors try to remain neutral on the interpretation of "moral judgement" as to whether it is to be understood in cognitivist or non-cognitivist spirit. However, a big part of the counterarguments target either the cognitivist or non-cognitivist version of the motivational internalist thesis. In this work, as is clear, we are mostly concerned with the cognitivist motivational internalism, which understands moral judgements to be truth-apt beliefs. From the perspective of moral psychology those judgements are usually understood to be beliefs.

Also, I believe that the following is indeed very relevant when thinking about the cognitivist motivational internalism. Zangwill insightfully specifies that "it is at least necessary that moral beliefs motivate". And adds to the internalist claim still another requirement of Kit Fine: "we should distinguish modality from essence (Fine 1994); for there can be necessary connections that are not essential connections. The internalist needs to claim not just that moral beliefs are necessarily motivating, but that motivation is essential to moral beliefs" (Zangwill 2008: 94). We will discuss this question later on some more, for now I just want to draw attention to the point that the thesis generally given as definition of motivational internalism might not be exhaustive. For the present purposes, though, it will suffice. Think of the charges with "moral fetishism" that certain MR_{MI} face in virtue of this relation being not an essential relation.

(2012) the position of 'simple internalism'[38] is expressed by the following claim: "necessarily, if a person judges that she morally ought to ϕ, then she is (at least somewhat) motivated to ϕ" (Björklund et al. 2012: 125). In another paper Strandberg gives a similar formulation of internalism: "It is necessary that if a person S judges that it is morally right for her to ϕ, then S is motivated to ϕ" adding that he takes it to mean that such a person S "is motivated to some extent to ϕ" (Strandberg 2012b: 27-28). The two formulations provided above may be treated as synonymous or not depending on whether one reads "it is morally right that she ϕs" as synonymous to "she morally ought to ϕ". I leave it open if it is necessary for an internalist to think that they are synonymous, however, I will treat them as such and will use the expressions interchangeably, unless noted otherwise[39].

Terminological issues: how to term the most general internalist claim. The term for this most general internalist claim in Björklund et al. (2012: 125) is "simple internalism". In the aforementioned paper by Strandberg (2012b) the (fully specified) internalist claim[40] is termed, though, "strong internalism" or "strong version of internalism". However, I argue that we should give up the practice of naming the distinctions in such relative, uninformative terms. Strength and simplicity require clarification: strong or simple in accordance with what criteria? Moreover, the basis on which the "strength" or "weakness" (or "modesty" or so) is assigned to views in this debate is not always the same.

[38] They are certainly not alone in defining it like this (Cholbi 2011; Zangwill 2008; etc.), their paper is rather giving a comprehensive review of the evolution of the debate and summarises the general tendencies, including the most popular definition.
[39] "Ought" by some is treated as indicating severe necessity: if you ought to do something, it is something that you must do, it is required that you do it. However, some, like professor of philosophy Michael Ridge, disagrees and claims that "ought" in the moral debates should get its usual English meaning of recommendation back.
[40] "It is conceptually necessary that, for any action ϕ and person S, if S judges that it is morally right for her to ϕ, then S is motivated to ϕ" (Strandberg 2012b: 28).

For example[41], Sigrun Svavarsdóttir thinks that the weak version of motivational internalism differs from the strong one only in that the former "does not take a stand on the mechanics of moral motivation – whether moral judgment motivates on its own or only in collaboration with some other mental states – but agrees with the strong motivational internalism that an agent has not made a genuine moral judgment unless he has the relevant motivation" (Svavarsdóttir 2006: 164). Whereas Mele assigns a label of a modest species of internalism to Dancy who holds that some beliefs are intrinsically motivating, but they are not essentially (necessarily) motivating (Mele 2003: 126). Hence, according to Mele's classification, the one belonging to a more modest, or weaker, version does take a stand on the "mechanics of moral motivation", but holds that the relation between the moral judgements and motivation is contingent.

Still another usage seems to be more common: Smith calls "weak" and Dreier describes as "modest"[42] those versions of internalism that are conditional, i.e. the ones that presume the necessary relation to hold not *simpliciter*, but under a certain condition, so, those which see the relation as defeasible. This seems to be Strandberg's usage as well, however, such authors as Lenman (1999) and Zangwill (2008) would consider his "strong internalism" to be a "weak" version of motivational internalism. E.g. Lenman claims to be following Brink in making the distinction between the strong and weak internalism: "Here we may follow Brink and distinguish strong internalism whereby to make a moral judgement suffices to motivate someone

[41] I will only give several examples of the different usage and will not aspire to give an exhaustive analysis of the possible meanings of "strong", "weak", "weaker", "weakest" and other predicates of internalist claims. I am therefore intentionally omitting, in this respect, discussion of Mason (2008), and many others.

[42] "But let us call modest internalism the principle that in normal contexts a person has some motivation to promote what he believes to be good. Modest internalism is not vacuous, though of course it is weaker than strong internalism" (Dreier 1990: 14).

to action from weak internalism whereby all that is necessary is that it provide *some* motivation" (Lenman 1999: 441; italics mine)[43].

Of course, one could argue that some of the usages of the "strong", "weak" (or "modest") are better than others presented here and choose one claim as basis for the label. For example, to choose the one which is the most common in the debate. However, I believe that instead of that, the various claims should preferably be termed with reference to such a difference between them that does not require any further introduction of criterion (such as criterion for robustness of the position) in order to point out its meaning to others than the author. Besides, there is still another basis for assigning "strength" and "weakness" to the formulations of internalism which I will take up in the upcoming pages.

The bases for aforementioned distinctions are different, but I would not endorse attribution of "strong" and "weak" to the term of "internalism" for making distinctions on any basis. As for my present interests, the distinction of the internalist claims that Smith, Dreier and Strandberg have in mind will be relevant, that is, what will matter in my text is the character of the necessary relation: whether it holds under a certain condition or unconditionally, *simpliciter*. Thus, the terms of "conditional" and "unconditional" internalism respectively. I believe that this terming expresses the wanted distinction the best. And though in (2013) Strandberg and Björklund have introduced the label of "generic internalism", I think it is not as successful as "unconditional internalism": "conditional" and "generic" do not have a clear, explicit relation. Therefore, I will use the unconditional/conditional distinction when necessary, instead of other terms which are more frequent in the debate, but which also remain vague and require references to specific authors.

[43] Though in "Moral Motivation" (1997) Brink gives still another meaning to the strong/weak distinction, whereby 'strong' denotes that the moral judgements entail motivation, whereas 'weak' means that they are only accompanied by motivation. Not seeking to give an exhaustive analysis about how the strong/weak distinction was and is employed in the debate, this suffices to make the point that these notions are ambiguous.

Primary analysis of the unconditional MI. The first thing to notice is that the claim in the debate has been understood first of all as an *a priori* necessary claim, and so the necessity of the relation between moral judgements and motivation is a *conceptual* necessity. However, there have been discussions about motivational internalism as a claim that the necessary relation is *a posteriori* necessary and about its philosophical relevance as well (Björklund et al. 2012: 32-34). However here, when not explicitly noted otherwise, the claim should be understood as having *a priori* status.

Second, in the definitions given above, the requirement on the "moral ought" is to *motivate to at least a certain extent*, "at least somewhat". However, this formulation, without further context, is ambiguous between two readings:

1) moral motivation does not need to be overriding with respect to other kinds of motivation relevant to a given agent faced with a practical decision;

2) the strength of moral motivation does not have to be proportionate to or commensurate with the strength of the moral judgement (which, e.g., can be said to be expressive of a belief that there is a normative reason to do something, or so).

First of all, we should make clear that usually what is meant here, is that the moral motivation does not have to be necessarily overriding. Strandberg clarifies the part of "is motivated to some extent to ϕ" by adding "not that she is most motivated to ϕ" (Strandberg 2012b: 28). Zangwill also first and foremost talks about this meaning ("some motivation … this motivation need not override other motivations") contrasting it to a stronger view, on which "the motivation that springs from the moral judgement necessarily overrides all other sources of motivation" (Zangwill 2008: 93-4).

However, it seems that even if many theorists do not spell it out explicitly, they do read the expression in the second sense as well: MI requires agents to have motivation of at least minimal strength – regardless of the

strength of the moral judgement[44]. And this, I will claim, is implausible. There is not too much discussion on whether we *should* understand that "some motivation" still falls under further restrictions on its strength, but it is a crucial question to settle. So I will discuss the implications, advantages and disadvantages of both possible answers to it straight away.

Criticism of the unrestricted MI. Without posing any further restriction on the strength of motivation, the motivational internalism (MI) thesis, whether in its conditional or unconditional guise, is rather weak. There we have another occasion to term MI "weak"; however, as I still think it useful to avoid this adjective altogether and specify the various claims with reference to the main difference in between them, I chose not to term it the "weak" version, but rather the "unrestricted" MI. The unrestricted MI, then, is threatened[45] only by (the possibility of) instances of *total lack of motivation* in people who sincerely make (moral) judgements. The weaker than appropriate motivation (e.g. one's moral judgement is overriding others with regard to the question of what to do, but one's moral motivation is not overriding in this regard) is not a counterexample to this version of the MI. That is why depression, addiction and similar mental conditions (as well as the famous amoralist example) pose a challenge to such internalism only presupposing that those conditions eliminate motivation *completely*. These cases, then, do serve as counterexamples to motivational internalism, as they mean that it is possible that somebody (sincerely) makes a moral judgement that one oneself morally ought to φ and still remains *absolutely* motivationally inert, or indifferent, in this respect.

[44] It is well seen when externalists criticise the unrestricted version of MI even in cases where an author, e.g., Smith, is defending a restricted version of MI. I will elaborate this point and explain the distinction of restricted/unrestricted versions of MI further on.

[45] *A priori* claims are, of course, not threatened by any counterexamples, however, in face of the counterexamples the claim – at least in its current formulation – loses its plausibility. In light of such examples, we may either question its status or its formulation, or its relevance to our actual world. Thus, the possible counterexamples are relevant and can be useful for perfecting the formulation of the MI and for providing an acceptable analysis of it (a plausible conception of moral judgement).

Rather often weak-will is omitted from the debate by the critics of the MI: the unrestricted kind of MI is not threatened by the weak-willed people as far as "weak will" does not imply *absence* of will[46]. Such an interpretation would certainly be acceptable to some theorists. For example, Svavarsdóttir refers to Michael Stocker's observation "that under conditions of deep depression, severe cases of weakness of will, and other maladies of the spirit, the connection between moral judgment and motivation is often broken" (Svavarsdóttir 1999: 163-164). I suppose that the need for "severe cases of weakness of will" shows that akrasia is taken not to eliminate moral motivation, except in its extreme manifestations. However, I am not to claim that all externalists suppose that akrasia implies there necessarily being some kind of motivation in accordance to every judgement involved in the decision. Some of them can well think it is on the same ground as the other mental conditions that eliminate moral motivation completely. But on the former understanding and on this, unrestricted, interpretation of the MI, weak-will (whatever it is) poses no problem for internalism, and the amoralist example is of the same kind as the cases of depression and the like[47].

However, I want to argue that this kind of reading (the unrestricted MI) is implausible for two main reasons:

1) trying to show it is not true, externalists oversimplify our mental life and suggest false phenomenology of it; in other words, this reading suggests an oversimplified picture of human mental life;

2) internalist accounts, based on it, are devoid of any explanatory power – not just the one they pretend to have, but of none whatsoever[48].

---

[46] That is how once Caj Strandberg explained to me the "weak-will". He thinks that internalist position allows for the akratic actions.

[47] Perhaps not of the same kind if we think that "amoralist" serves as an *a priori*, conceptual counter-example to MI, and "depression" and the like refer to something more than just conceptual possibilities. The status of the latter, however, is not so clear. But we will come back to this topic.

[48] This is only true of the unrestricted version of MI, though, the restricted one, as we will see, does not share these flaws, quite to the contrary.

So the unrestricted version is useful neither for internalists, nor for externalists, both should aim at defence or criticism of the restricted version. Let me explicate the flaws of the unrestricted MI in the order I presented them.

(1) First, one can only agree with Zangwill who says that "it is not the case that either we believe something or we don't or that either we desire something or we don't. Beliefs come in degrees and desires come in strengths. … Our mental world is not black and white" (Zangwill 2008: 95). We talk about weighing reasons and about stronger and weaker desires that we have, even of desires that are too weak (to lead to action) and too strong (to resist). Therefore, the differing strength of these should be accounted for in a proper meta-ethical theory. If beliefs were not differing in strengths, as well as desires, much of our moral debate would become futile. Of course, one cannot say that the theorists I refer to do not admit that we can talk about differing strengths of moral reasons, but their counterexamples rely on the cases where moral motivation is totally wiped out. As Zangwill claims, such simplifications handicap the entire debate (ibid). He addresses this requirement to the adherents of the MI, but I think it goes for both sides: neither internalists, nor externalists should oversimplify our mental life.

This criticism applies especially to the various purported counterexamples, say, that of depression. Even though it is hard to understand what the status of such a condition is (does it have anything to do with the depression some actual people have[49] or it is just one of the avatars of the amoralist thought experiment?), it seems quite implausible to think that people,

[49] I see this question especially pressing when we discuss results of experimental philosophy (see, for example, Strandberg and Björklund 2013). When people's intuitions are being tested, do they understand "depression", "psychopathy" and the like in accordance with their perfect or imperfect knowledge or actual experience of these conditions or with their imagining what those conditions consist in or with the uninformative definition that the experiment conducing persons present in the description of a particular case? It is also interesting whether these different sources of knowledge are not in tension and if they do not influence these people's answers? But these are questions to be discussed elsewhere. I can only say that there are different positions on the status of these conditions in the literature that this debate consists of.

struck by hardships, suddenly begin having *no moral motivation whatsoever*[50], phenomenologically – start feeling *absolutely no inner conflict*[51].

But let us grant that this is plausible as it is conceivable. In any case, allowing for this reading of MI, we lose many other cases that could have relevance in shaping the answer to the problem of the relation between moral judgement and motivation. These are the cases of addictions, compulsions, akrasia and other states that are more common to our everyday life, in which not only we perceive the reasons of distinct strengths that we have and that put claims on us (we evaluate them to be distinctly important), but also feel torn apart, I suppose, by distinct motivations (of different strengths).

But of course, one can object that phenomenologically the difference between a very weak motivation and none at all is too slight to be noticeable. It can be equally weird to say that a person is motivated to some ridiculously small extent (because some theory supposes there has to be at least some motivation present), even if she does not feel anything. It is not and should not be a matter of phenomenology (it does not dictate a theoretical position so directly), but of its theoretical relevance in action explanation. And, in this respect, does it make sense to posit any such small motivation? To answer it, let us move to the next point. However, by this first point I meant to say that it is not necessary for externalists to restrict the extent of their counterexamples by such artificial means and by sometimes proposing a false or too rough a picture of how human motivations function. They can have and use both the examples that show the insignificance of motivation and the absence of it.

---

[50] As Cholbi claims (2011), it is quite to the contrary: the actually depressed persons usually do not lack in moral, but rather in prudential, or self-regarding, motivation. Though, as far as I know, Cholbi does not specify if the prudential motivation is totally absent, or is just insignificant enough to issue in action (nor whether they can be said to make self-regarding judgements at all).
However, I have already noted that the status of "depression" in the counter-examples is not clear, i.e., if it has anything to do with the depression as a clinical condition.

[51] Maybe externalists here could say that the apparent tension is rather between something else than the two different desires; for example, between volition and desire (conflicting elements of different levels) or between beliefs (cognitive conflict) or so? I am not sure.

They can, because externalists can acknowledge that even there being some kind of relevant motivation does not by itself mean that that motivation arises *necessarily* or *because of* the judgement made, or, to use psychological vocabulary, the belief had.

(2) As to the second point, first of all, I want to remind that what we are talking about is a motivational internalist thesis which does not require the moral motivation to be necessarily overriding. Second, motivational internalism should first of all be thought of as a more general position – that the motivation to act in accordance with one's *normative* judgement is necessary and *essential*. It is hard to think of a contemporary moral motivational internalist who thinks that morality is the only source of normative reasons, however these are defined. Or of an externalist who thinks that only moral judgements do not necessarily and essentially issue in motivation, whereas other kinds of normative judgements do. I believe that usually what distinguishes the two camps is exactly this more basic matter: distinct positions on whether one's normative judgements motivate necessarily and essentially. Only after this is settled, we can proceed to the discussion of those internalist theorists that claim the superiority of the moral judgements over other kinds of normative judgements, or those who claim that morality is not a matter of rational requirements, and so on. I will treat the matter this way.

In order to make my second point, I will give an example of decision making, consistent with the unrestricted version of MI. Let us say, that a certain person Patricia judges that it is morally right for her to take her ill friend Anna to a hospital circumstances being such that Anna has nobody else to help her out at that moment and feels terrible. Patricia judges this stronger than she judges that she ought to meet Jim for a romantic dinner as they are in love and she has promised to come that evening, where the latter action is non-moral (but not necessarily immoral). Then, if she is motivated to help Anna more strongly than she is motivated to meet Jim, all things considered, her moral judgement issues in relevant motivation. However, if her motivation to meet Jim is stronger than her motivation to help Anna, whereas she judges

more strongly that she ought to help Anna, but is motivated more strongly to meet Jim, her stronger judgement does not match her stronger motivation, she is weak-willed or so. And, alternatively, if Patricia judges that she ought to meet Jim in these circumstances more strongly than that she morally ought to help Anna, and has stronger moral motivation to help Anna than the alternative motivation to meet Jim, her overriding motivation and her normative judgement all things considered differ in content; but if she is motivated more strongly to meet Jim than she is motivated to help Anna, her non-morally normative judgement issues in relevant motivation.

Here we can see that all those four cases satisfy the unrestricted MI requirement, because all the normative judgements, naturally, moral including, issue in at least some motivation, of at least the smallest strength. However, if the requirements of internalism are so low, and the strengths of the motivations, issuing from judgements, are contingent, then an internalist cannot explain why a person, who judges one thing stronger than another, is still motivated more to do something else than her strongest normative judgement says to do[52] – an overall motivation does not follow an overall decision. Though the relation between the *pro tanto* judgements and relevant motivations is necessary, at the *all things considered* level, it seems to be totally accidental, or at least unpredictable, suggesting that internalism holds on the first level, but not on the second. Such unrestricted MI would be compatible with the existence of depressives and such, understood as the ones who do not lack moral motivation completely, but which have not enough of it.

As Mason notes considering similar matters, the only difference between such form of internalism and externalism is "that weakest internalism says that when there is a moral judgement there is always some level of motivation, however slight and ineffective. On this picture, the strength of the motivation that is necessarily attached to the judgement is random — it could be anything

[52] Of course, proponents of *conditional* MI can say that those people do not comply to some additional condition, but if the condition was added to the making of judgement as one more necessary condition, and we got the same result, they would not have how to explain it. The problem persists.

from the tiniest speck of motivation to motivation all the way to action, and the strength of the motivation is not tied to the strength of the reason that is judged to apply" (Mason 2008: 144). The question that arises is why to presuppose that there is this necessary and essential relation, if it does not play any role whatsoever?

Presupposing such a thing without the restriction I hint at would really be futile in quite some cases. I agree that such presupposition should not be the core of the MI as it does not have any explanatory power. However, the restricted MI would presuppose such a minimal necessary motivation in cases, where the relevant judgements are weak to the same extent, but that, as we will see, is not the same thing.

Internalism claims to be able to (whereas externalism cannot) explain why changes in one's behaviour reliably follow changes in one's judgements and why agent's values (for which she has direct concern) explain her actions. Also it supposedly can make sense of our intuitions that the words of a sincere person are consistent with her actions. However, internalism as represented above cannot explain any of these things. Whether these things happen is just accidental. The claim that there is a necessary and essential relation between the normative judgements and motivation to act accordingly is absolutely futile on the all-things-considered level. Even if the unrestricted MI is correct, we cannot be sure, that a sincere person who judges strongly that one oneself morally ought to help a friend and who judges less strongly that one (non-morally) ought to go on a date, will have a stronger motivation to help a friend than to go on a date. Therefore, the difference in between such an unrestricted version of MI and externalism is insignificant, internalism having no explanatory supremacy. Indeed, as Mason points out, the explanatory role of such motivation is limited and the claim serves only to "satisfy the basic internalist intuition that it is odd to judge that you ought to do something and yet not be motivated at all. But without an independent argument for internalism, that intuition is not a good enough justification for adding the internalist clause to the theory" (Mason 2008: 144).

Thus, it seems that internalism is an expression of, or a theoretical explanation of, the reliable relation between a moral judgement and moral motivation – that is why it is needed. But if it cannot explain that relation, there is no good reason we should defend it. In other words, it seems that people presuppose it because of its prognostic and diagnostic powers. But if internalism fails, then externalism is even better equipped to explain the relation of judgements and motivation. While internalists would stumble (why would one be less motivated to do what one judges more strongly that one should do, if the relation between a moral judgement and motivation is essential), externalists could explain it: because the relation is not essential. They could, e.g., say that people's motivations depend on their predispositions or desires, and not on (the strength of) the relevant judgements.

If an internalist advocate of an unrestricted MI is to say that weak-will is an explanation of a differing motivation, I would disagree: it is a name for a situation that itself requires explaining. Such explanations can be various, of course, such as, I suppose, "because one is severely depressed" or "because this is just a singular action upon which not much hinges, and so one gave into temptation" (or less sophisticated), and so on. "Weak-will" is not the final explanation, it is barely a statement of inadequacy (or incoherence) of one's best decision and one's strongest motivation for action. This statement can also be expressed in other terms, or so I will try to show later on and that observation lets me suspect that "weak-will" can be just one possible characterisation of the incoherence state among others which fall under a wider term and form a category.

The restricted version of MI, though, is different: it can make sense of the aforementioned intuitions and phenomena. It requires not only that the motivation necessarily follows the relevant normative judgement, but also that the strength of that motivation is proportionate to the strength of the judgement. However, it seems that on such a conception of MI, weak-will (and other similar instances of incoherence) is impossible. The only way to account for it is to accuse the person of insincerity or of linguistic incompetence. But

we will see that these are flaws pertaining to the unconditional MI, but not to any restricted MI position. Say, a *conditional* version of restricted MI will be able to cope with it in more plausible ways.

The point to be established from this discussion is only that the unrestricted version of MI, whether unconditional or conditional, is unacceptable. It makes us disregard a whole lot of instances relevant when trying to define MI as well as possible, to make the definition more refined, and so it makes us consider as counterexamples only the thought experimental cases. Besides, because of the MI being too vaguely defined, it can be neither confirmed, nor disconfirmed[53], as well as it cannot explain the main things it is meant to explain. Therefore, a more refined version of MI is needed, and it can be achieved first of all by adding a restriction on the quantitative dimension of the judgements and of their corresponding motivation.

The proportionality, or commensurateness, requirement. The question is what such a requirement should look like, and whether we already have any examples of it, and, if not, what its introduction would mean.

Zangwill calls such requirement the "Proportional Determination Thesis": "The degree of a person's moral belief that he ought to do something proportionately determines the strength of his desire to do it" (Zangwill 2008: 95). The "determination" here is to account for the necessity of the relation of the quantitative dimension of appropriate belief and desire.

A clear adherent to the idea that a restriction of the MI claim is needed is Smith. He hasn't given an explicit formulation of the requirement, but talked about it as of one more requirement of rationality. For example, Smith in (Smith 2001) aims at the defence of the claim that even if the belief that one has a reason for doing some particular action is false, one's having a relevant desire still *makes sense* while one's believing the aforementioned thing and not

[53] Whether a restricted version can in principle be confirmed or disconfirmed depends on what it is like and on its status (an *a priori* or an empirical claim). However, it is not true of the unrestricted version.

having the appropriate desire or being averse to that action does not make sense. Then, he adds a refinement: "It seems to me that we can draw the even more fine-grained conclusion that *S's* having a desire of a certain strength to do *x* in *C*, when he has the belief, true or false, that if he had a maximally informed and coherent and unified desire set then he would have a desire of that strength that he does *x* in *C*, makes sense in a way in which his having a desire of some alternative strength to do *x* in *C* simply doesn't" (Smith 2001: 259-60, n. 2).

From what was cited, it seems that Zangwill by a "degree of belief" refers to something else than Smith, i.e. perhaps according to Zangwill, it is a measure of how certain the deliberator herself is that her belief is true. For Smith, though, the strength of the relevant desire, which embodies motivation, is tied to the strength regarding the content of the belief – to the strength of a hypothetical desire (what the deliberator would desire to do if she was fully rational). Strength of the judgement, it seems, for Smith depends on how justified the belief is – how strong the reason, grounding the decision of what is desirable, is. A similar idea that the relevant dimension to determining the strength of the belief is that of the reasons has crossed the mind of at least one critic of internalism: Mason[54].

However, in (2002b) Smith adds two more factors that he holds relevant to determining strength of the due motivation. He lists three features of evaluative judgements and calls them *certitude, robustness* and *importance*. The latter is what I mentioned just before and Smith holds it to be a feature of evaluative judgements in particular (non-evaluative ones do not have it). Importance indicates how desirable a person judges something to be; it is fixed from the perspective of the omniscient, or abstracting from certitude. Certitude shows the level of confidence that the person has in what she judges to be the case. It can be measured abstracting from importance, by "how much they

[54] "If the moral ought is not overriding, the formulation of weak internalism will be a bit more complex. The appropriate claim would be that the strength of the motivation should be commensurate with the strength of the reason" (Mason 2008: 144, n. 11).

[people holding certain evaluative beliefs] would be willing to bet on one outcome as opposed to another under circumstances of forced choice" (Smith 2002b: 307). Robustness shows how stable a person's confidence is in what she judges to be the case, in face of the incoming information and reflection. It is fixed abstracting away from certitude and importance, and measured in accordance with how much a subject would be willing to bet on some outcome as opposed to another over time. So while certitude measures the levels of confidence in one's judgements synchronically, robustness does so diachronically.

So, according to a cognitivist Smith, evaluative judgements which consist in beliefs, have these three features that are all relevant to the strength of the relevant motivation, which consists in desires. Smith claims to give analysis of the common-sense understanding of what features evaluative judgements have. I believe that the justifiability of such judgements from the first personal point is the most intuitively plausible dimension that should be mirrored in the relevant desires. Then, sometimes, certitude of the judgements comes into our considerations. However, as far as common sense is concerned, robustness perhaps is not something that we think of. But whether Smith's calculus of strength is complete or not, plausible or not, is another and, I think, not the most important question. Wherever evaluative belief gets one's quantitative dimension from, the idea behind the requirement is to relate it by necessity and proportionately to the quantity of the motivation-encompassing desire. So a more general (if we think it applies not only to moral judgements, but to evaluative judgements in general, as above) proportionality, or commensurateness, requirement would perhaps be such:

The strength of the evaluative judgement has necessarily to be proportionate to, or commensurate with, the strength of the motivation to act accordingly[55].

[55] Here, so far, I put the essence problems aside. Also, I suppose that evaluative judgements are understood as having a practical upshot.

Implications of this restriction are such that now it can be more easily targeted and the counterarguments to the restricted versions of the MI can be categorised differently and they posit slightly different threats.

Whereas to the unrestricted MI weakness of will posited no challenge, now it is one of the many cases that the formulation of the MI has to account for. So weakness of will is now on the same footing as the cases of addiction, depression and the like, taken not to necessarily issue in total absence of the moral motivation[56]. However, the counterarguments, based on a supposition that the *complete* indifference to the moral judgements is possible, such as the amoralist case and, possibly, others, would require of the restricted version of MI a different kind of answer: internalists' answers to the two groups of counterexamples should be diversified.

One more thing to notice is that the restriction enables to bridge the crack between the *pro tanto* and the *all things considered* levels: the restricted version of the MI that does not require a necessarily overriding moral

[56] We will see later on that it is not always obvious how to treat examples of "depression", "listlessness" or such, i.e., as hypothetical cases or as the clinical cases. Here I give "depression" a more mundane reading, as I think that no aim is served by giving more names to the example of the amoralist, where it refers to a being with a complete absence of moral motivation, and which serves as an *a priori* counterargument to the MI. I see that the "amoralist" can have negative connotations and be read differently, though. I.e., the difference between an amoralist and a depressed person may be understood more or less as difference between a cynic *vs.* a psychologically handicapped person (as worthy of our sympathy, being unhappy about one's failing to be motivated, not being indifferent wilfully, or so). Perhaps, but should there be a whole host of such notions employed to account for a single counterargument? I think that manoeuvre just brings more confusion of terms.
Another point worth noticing: it is not obvious if "weakness of will" cannot be a category, comprising all those more specified cases of the mismatch between one's best judgement and one's strongest motivation. Or should it be viewed more traditionally – as something that does not have a pathological cause? Though the questions of addiction are also not that easily classified; there is literature denying addicts the status of victims and rather finding causes of addictions in conscious choices or just in giving in to temptations.
Besides, it seems that in cases of akrasia the "best" judgement can be the only normative judgement that the person makes, but the overriding motivation can be such which does not come from any normative judgement at all. Can the same be said about an addict's actions or do we need to admit that the lesser normative reason is that of pleasure? It is not obvious for me how to finally define akrasia and whether these other cases are just instances of it or not.

motivation, requires that the strongest practical judgement issues in the strongest motivation.

This restriction should be understood as applying both to the more basic internalist requirement on practical judgements (or evaluative judgements, as in Smith's 2002b) and to the moral MI more specifically, and both to the unconditional and conditional versions of the moral MI.

The restricted MI is also able to explain why the changes in motivation reliably track changes in judgements, and that people act in accordance with their values, the relation between one's decisions and actions. And Michael Smith is certainly an advocate of one of the most prominent accounts of the restricted (conditional) MI.

One more question that can arise in face of the proposition of this restriction is not just that about the calculus of the strength, but also whether "proportionality", or "commensurateness", requires the strengths of a judgement and relevant motivation to be exactly of the same extent. Perhaps the strengths could be highly similar, even if not exactly the same? I suppose, the logical answer would be that they should be perfectly identical, but this is not needed for the restricted MI to hold and retain its explanatory power. Thus we could allow that the strengths should be at least highly similar, but that the difference between them should not exceed the difference in strength between any alternative (and relevant to the choice) judgement and relevant motivation. Such would by my very rough restriction on the possible imperfections of strength equality.

2. Introducing the rationalist internalism (RI)

The "indifference argument". Zangwill believes that the "indifference argument" is *the master* argument against internalism (and in favour of externalism). The argument starts from a whole bunch of counter-cases to the MI which, according to Zangwill, can be collected under the label of "moral indifference". He points out in (Zangwill 2008: 101) that Foot was the first in

the debate to draw attention to indifference (1978), then Michael Stocker appealed to indifference (1979), David Brink – to amoralism (1989), Al Mele – to listlessness (1996) and Svavarsdóttir – to cynicism (1999). In view of these cases it is claimed that externalism can explain the cases of indifference best.

This argument, according to Zangwill, "involves an appeal to the possibility and actuality of a certain kind of indifference to moral considerations" (Zangwill 2008: 92). Zangwill's way of putting the counterargument is much stronger than the previous ones, because he appeals to more mundane cases of indifference rather than to cases which are highly controversial. For example, he acknowledges that it is indeed disputable if an amoralist, i.e. that who does not care at all, is possible or actual, as well as cases where a person ceases to care at all (at least they are not common, "an inductively weak basis for a general claim" (Zangwill 2008: 106)). However, the cases of people caring less than they did before ("trans-temporal cases") or cases when some people care less than others ("trans-personal cases") are actual.

What needs an internalist's explanation, then, is the (interpersonal or intrapersonal trans-temporal) variation in strength of desire while the degree of (moral) belief stays constant. One has to acknowledge that it may be hard to measure the degrees and strength of desires and beliefs between different people, but it is rather plausible to talk of the differing strength of desire in the same person trough the time. So it is a real problem even for the restricted version of the MI. According to Zangwill, though, the best explanation is delivered by externalism: a moral belief and a motivating desire are two distinct entities, thus the variations in one while the other remains constant.

Ways to deal with the counterexamples. One of the ways to deal with the counterexamples, or the indifference argument, and the one which proved to be the most ineffective, is to say that all of those, who make moral judgements and are still motivationally completely indifferent, are not *really* making them. They are only making them in some kind of "inverted commas"

sense, therefore, they either lack the linguistic competence or just report others' judgements.

Another path to take is to question the very possibility of amoralists and the presumption that depression and other such mental conditions prevent motivation completely. Here one can take a stance either while still remaining in the armchair or by putting some relevant empirical evidence on the table.

The unconditional MI (both in its restricted and unrestricted forms) can only deny the counter-examples: claim that these cases are either impossible or that the people in the cases lack linguistic competence. Of course, many, even within the camp of internalism, found the way of denial an improper answer and acknowledged the possibility of somebody's being unmotivated by one's moral judgement. So the third way open to such theorists was to introduce into the unconditional internalist claim a proviso that could accommodate the aforementioned cases in some way or another. Advocates of this latter strategy have proposed versions of *conditional* motivational internalism stating that the conceptually necessary relation between the moral judgement and the relevant motivation holds under a certain condition. Note though, that for an adherent of this third strategy the other two ways for dealing with some of the counterexamples that are available to the unconditional MI are available as well. That is why we may need to look into some such manoeuvres. But we will not investigate the history and failures of the first two strategies in depth, rather, we will be selective. I only want to say that even the conditional MI needs to treat the counter-argument of amoralist and the cases presented by Zangwill in different manner as the first one implies the absence of any moral motivation, whereas the latter only suppose the weaker than appropriate motivation.

In contemporary meta-ethics the unconditional MI thesis is mostly a nonstarter: "[i]n contemporary metaethics, it is regularly assumed that this view is too strong, since it seems possible to conceive of someone who makes a moral judgement but fails to be motivated accordingly because she suffers from, e.g. apathy, depression, exhaustion, or emotional disturbance"

(Björklund et al. 2012: 126). And the first two strategies are not very successful even by some of the internalists' eyes. So let us begin the analysis of the conditional MI.

To sum up, whereas the unconditional internalism claims a conceptually necessary relation to hold between a moral judgement and the relevant motivation unconditionally, introduction of some proviso means that the relation holds necessarily only upon a certain condition. That condition can be spelled out in several ways. I will give some most prominent examples.

One of the varieties of conditional MI is termed "communal internalism" and, for example, its advocate Jon Tresan defines it as a thesis "that the internalist necessity obtains at the level of communities rather than individuals" (Tresan 2009: 180). This condition allows for amoralists and other cases of moral indifference at the individual level, but not at a communal level which is supposed to be the source of our internalist intuition. It makes sense to think that the moral practices persist even if some of the individuals do not participate in them: "moral beliefs require the characteristic moral practices of socialization, norm-enforcement, and self-guidance, but once such practices are up and running, moral beliefs may be acquired by individuals who do not themselves participate in the practices" (ibid.).

Another example is of internalism which defines the condition as "normal circumstances": "in normal contexts a person has some motivation to promote what he believes to be good" (Dreier 1990: 14). However, a good and not *ad hoc* analysis of the normality is not given ("Though I think I have successfully argued that we do have a grip on the conception I need, I cannot now provide an analysis" (ibid.)).

Björklund and others (2012) also describe MI which relies on the psychological normality of the moral agents and an MI which requires moral perceptiveness from the moral agents. However, all of these versions of MI suffer from problems. For example, the communal MI seems not to satisfy the pre-theoretical intuition that the necessary relation holds for every individual. The normality condition seems not to allow of a plausible, non *ad hoc* analysis,

besides, not all of the counter-examples to MI involve people who are abnormal in psychological or any other way. The moral perception comes from realist$_{MI}$ theories that are problematic in their own way.

However, there is one more popular proviso, which is the most promising by my own and some others' view[57], and it is formulated in terms of agent's rationality: if a person judges that it is right for her to ϕ, then she is motivated to ϕ or is practically irrational. To put it otherwise, Strandberg's formulation of the rationalist internalist position is as follows: "It is conceptually necessary that, for any action ϕ and any rational person S, if S judges that it is morally right for her to ϕ, then S is motivated to ϕ" (Strandberg 2012b: 30).

The RI reply to the amoralist challenge. Before delving into the analysis of the RI, let us explore its answer to the amoralist challenge. As mentioned, the RI can answer the indifference argument by introducing a proviso into the MI whereby the cases of motivational indifference are to be classified as failure to comply with the condition or a breach of this additional requirement. However, the cases of *complete* indifference need to be dealt with differently: the very possibility of there being cases that deny what is at the heart of the MI, i.e. the necessity of relation between moral judgements and relevant motivation, deserve exceptional attention.

Smith reacts to the externalist challenge slightly differently than those saying that amoralists report judgements of others. Smith holds the defenders of internalism to be right in claiming that amoralists do not really make moral judgements, or that they use moral concepts in some kind of inverted commas sense. He suggests that the debate on whether motivation is a necessary condition or is rather optional, an extra, for mastery of moral terms and ability

[57] For example, Strandberg claims: "The most promising version of weak internalism is what I refer to as 'rationalist internalism'" (Strandberg 2012b: 26). As already noted, for him "weak internalism" refers to conditional MI, or the version of internalism which proposes that the necessary relation "holds only for those who satisfy a certain condition" (Strandberg 2012b: 25), that condition being unspecified. And theories specifying the condition in terms of agent's rationality get the label of "rationalist internalism".

to make moral judgements, has the same structure as the debate over the conditions for mastery of colour terms. On the one side there are those who say that in order to be able to make colour judgements one needs to have appropriate visual experience or moral motivation respectively. On the other side we have those who hold that "the ability to use a term whose use is reliably explained by the relevant properties of objects is enough to credit her with … the ability really to make colour judgements (moral judgements)" (Smith 1994: 70).

However, in such debates one needs an independent reason to determine which side is right as both assume what they try to prove. And according to Smith, we can get such an independent reason in virtue of differing potential of the two theories to explain why a change in motivation reliably follows a change in moral judgement ("at least in the good and strong-willed person). To the contrary of Zangwill's argument, Smith believes that it is internalism and not externalism that explains this phenomenon best.

Internalists cognitivists (I skip the non-cognitivists) believe that this connection is to be explained internally: "it follows directly from the content of moral judgement itself" (Smith 1994: 72). Externalists explain it externally, that is, as following from the content of motivational dispositions of such good persons. As for the stipulation about the relation holding in good and strong-willed persons, both have a story to tell as to what counts as a good person. If being a good person in one way or another explains the reliability of the relation between the moral judgements and motivation, then for externalists being a good person will mean having a disposition to moral motivation. In other words, a good person for them will be such as to be willing to do the right thing. For the internalists good person will rather be that one who will non-derivatively care about doing the right thing, i.e. who wants to do what she judges to be the right thing to do, where this is read *de re*, and not *de dicto* (Smith 1994: 73).

But we are already familiar with the argument of moral fetishism, so I will not repeat myself. However, I bring into attention that Smith holds these

considerations to present us with the wanted independent reason for accepting either of the positions on mastery of moral terms. And in the face of externalism facing the charge of moral fetishism, we have a reason to accept an internalist position. Thus, it is preferable to claim that in order to make a moral judgement one needs to be motivated to act accordingly. Therefore, an amoralist does not present a genuine challenge to the MI nor the RI because an amoralist does not really make moral judgements and does not possess mastery of moral terms: in Smiths words, s/he tries to make it, but fails.

The rationalist internalist position. But what is there to this condition, why would rational persons, unlike, say, irrational persons, be necessarily motivated to do what they deem to be moral? Michael Smith, one of the most distinguished adherents of the rationalist internalism, calls what Strandberg termed "rationalist internalism" the "practicality requirement on moral judgement" and claims that another internalist thesis ("rationalism") explains it (together with one other claim). The explanation is that moral judgements are judgements that there are normative reasons for oneself to do the things in question: "If it is right for agents to φ in circumstances C, then there is a reason for those agents to ϕ in C" (Smith 1994: 62). Another missing link is supplied by what Smith thinks to be a platitude: "an agent has a reason to act in a certain way just in case she would be motivated to act in that way if she were rational" (ibid). These two claims in conjunction then allow to state that "an agent who judges herself to have a reason to act in a certain way – who judges that she would be so motivated if she were rational – is practically irrational if she is not motivated to act accordingly. For if she is not motivated accordingly then she fails to be rational by her own lights" (ibid).

However, Smith's early terminology may sometimes confuse his readers. First of all, it is so because intuitively "reason" is an objective term (at least in one of its senses, as far as it has the justificatory dimension). The formulation "If it is right for agents to ϕ in circumstances C, then there is a reason for those agents to ϕ in C" (Smith 1994: 62) may imply that the reasons that are had in

mind are objective entities, but we know that it is not necessarily true that if a person judges some action right, then there is for her such an objective reason to act in that certain way: one can be mistaken. So on the one hand, it seems that reasons can be "there" and apply to the agent, but she may be unaware of them and judge that something totally else is right, something that, in fact, is not a reason at all. On the other hand, only that which the agent is aware of, can motivate her. So at least in some cases, that which motivates us (what we judge to be right to do), is not a reason, and the real reason neither appears in the contents of the judgement, nor issues in motivation. That is why the second proposition, the supposed platitude, seems to employ a different, more subjective, sense of "reason".

But this confusion is due to the formulations that Smith presented in (1994) and which he clarified in further pages of his book and in his later publications. His conception is better conveyed in reformulation of his claims. In order to state the position of his and of other rationalist internalists explicitly in an elegant form, we may borrow Strandberg's flawless logical reconstruction of the rationalist internalist (RI) argument from (2012b: 30-31):

(1) *Rationalism*: It is conceptually necessary that, for any action ϕ and any person S, if S judges that it is morally right for her to ϕ, then S judges that she has a normative reason[58] to ϕ.

[58] First of all, I emphasise that the internalism/externalism debate primarily concerns the motivational potential of normative reasons, not that of motivating reasons which is obviously there according to the definition. Thus, "reason", unless noted otherwise, should be understood as a normative reason (I will talk about motivational reasons further on in the text). Second, rationalists define reasons in terms of rationality, even if there are attempts to defend the rationalist internalist thesis without adhering to the definition of reasons in terms of rationality, as well as without endorsing the whole argument (e.g., John Broome holds the rationalist internalist thesis to be correct in virtue of the principle of enkrasia). Only with this in mind, can the first claim be termed "rationalism" in a more familiar way. I maintain that rationalism is in essence and from tradition, a view on the nature of morality: that moral truths are knowable by reason alone; thereof the content analysis of the moral judgement. However, I hold that acknowledging adherence to rationalism in moral philosophy, one remains silent on whether there are other (than moral) kinds of truths that can be determined by reason alone and on their strength. That is why I believe that to call oneself a rationalist implies only subscribing to the idea of moral normative reasons (and other

(2) *Normative internalism:* It is conceptually necessary that, for any action ϕ and any rational person S, if S judges that she has a normative reason to ϕ, then S is motivated to ϕ.

(3) *Rationalist internalism:* It is conceptually necessary that, for any action ϕ and any rational person S, if S judges that it is morally right for her to ϕ, then S is motivated to ϕ.

The first and the second premises in the argument, however, employ a notion that is ambiguous: should "normative reason" be read in a *pro tanto* or in an *all things considered* sense? My view is that there could well be two versions of the RI in virtue of the different meanings of "normative reason", subject to different kinds of criticisms. Therefore, one should be careful to specify which version one is discussing (defending or criticising).

I take it that rather often the rationalist internalist position is by default understood to necessarily require the prevalence of the moral, thus, the argument is read in the *all things considered* sense[59]. For example, Joshua Gert, intending to restitute the true reading of Michael Smith's "reason", still falls prey to it: "It is possible to read much of what Michael Smith has written and come away with the firm conviction that he means to ally himself with the traditional moral rationalists, and that he holds that moral requirements are rational requirements" (Gert 2008: 1). But Smith *does* hold that moral requirements are rational requirements. The widespread misinterpretation of Smith is rather due to the default reading of "reason" as an *all things considered reason*, and therefrom thinking that what is required is required *all*

---

normative reasons) being determined by reason, but it does not imply subscribing to either *pro tanto* or *all things considered* reading of "reason".
Besides, defining reasons in terms of rationality does not compel to define rationality as responsiveness to reasons. I only highlight this here and discuss it elsewhere in this dissertation.

[59] That may be due to the philosophical tradition where rationalism is mostly associated with Kant. Also, perhaps it is because of the belief that the promises of internalism have to meet very high criteria? After all, the internalists claim to be able to explain why we expect a sincere person to act in accordance with her moral judgement (and so adding the *ceteris paribus* condition is much of a disappointment or acknowledgement of defeat?). I leave it unresolved.

things considered as well[60]. Whereas Smith claims that "rationalism might now be taken to be … the claim that our concept of moral requirement is the concept of ***a*** reason for action; ***a*** requirement of rationality or reason" (Smith 1994: 64-5; emphasis mine).

So because of the two possible meanings of "normative reason" we have two variants of the RI, or four theses (two variants of (1) and of (2)) that seem to be not equally intuitive and not equally plausible to the critics. Their plausibility will be examined in turn.

First of all, if we consider the first reading of the "normative reason" to be a "*pro tanto* normative reason", this reading makes moral reasons into an "unprivileged" subset of normative reasons. Some of the critics of the RI, to take just Strandberg and Gert, do not see any problem with acknowledging moral judgments to be normative reason giving considerations, contributing to the overall rational status of the action. However, this reading supposedly generates problems in the second, the normative internalist, claim. There are two main lines of criticism concerning claim (2). One of them is recurrent in many works of the critics of internalism: the notion of rationality cannot secure the necessary relation between every normative reason that an agent has and motivation to act accordingly. The other critical point is advocated just by several theorists, those who make a more refined distinction between the *rationally permissible* and *rationally required*.

However, the two critics just mentioned are very much against the *all things considered* reading of claim (1), which endows moral reasons with status of necessarily overriding reasons. Against this Strandberg claims: "It seems quite evident that a person may think it is morally right for her to ϕ but,

[60] Gert is preoccupied with refining the picture with the "permissible/required" distinction; he claims that some reasons rationally justify (permit), but not require certain actions, whereas others – not only justify, but also rationally require. Smith, however, does not make the distinction explicitly, but his "rationally required" or "a requirement of rationality" in the *pro tanto* sense in certain cases to a certain extent makes up for the lack of "rationally permissible" and I discuss this in detail further on in the text. So Gert in (2008), before proceeding to the criticism of Smith's account, is trying to do justice to Smith's "reason" restituting it the intended, but often missed *pro tanto* sense.

at the same time, think that she has a stronger reason to ψ" (Strandberg 2012b: 41). Gert is criticising such Kantian line of treating moral reasons as well in (2004: 13-14) and hereafter, also in (2008: 6). But on this reading claim (2), stating that a rational person is necessarily motivated to act in accordance with one's own normative all things considered reason, becomes *prima facie* more acceptable than on the alternative reading. Understandably, an externalist would not accept it in any case.

There is a line of criticism, recurrent in several works of the critics of RI, which threatens claim (3), i.e. the RI thesis itself. It asserts the condition of rationality to be insufficient to preclude some of the counterexamples to the motivational internalism.

So we will have to consider all four main lines of criticism, which will enable us to understand the RI better. I will argue that all the criticisms fail. Let me explicate the criticisms and the (pre)suppositions they rest on, as well as answers to them.

3. Conception of rationality

I will begin with the most general claim (3) so as to lay out at the very beginning the conception of rationality of the RI which is, naturally, the core of this position. Then, I will proceed to the defence of the claim (2) in its *pro tanto* sense (two lines of criticism to be overturned here), and finally, I will tackle the question of the superiority of morality, claim (1) in its *all things considered* reading.

3.1. What rationality is for the RI

Criticism of claim (3). This critical point is rather pervasive in the literature, however, I will show that it is grounded on a wrong presupposition. As the RI can easily deal with the counterarguments of accidie, depression and such, attributing or equating motivational indifference to the irrationality of the

deliberators, critics have to target (ir)rationality itself. The criticism mainly is such that one or another conception of rationality is not able to secure the necessary relation between a moral judgement and the relevant motivation that the RI is after.

Let me lay out the usual strategy that the critics of the RI employ, and only then proceed to particular persons. A critic selects a conception of rationality and applies it to some cases. The selected conception then proves to be not apt enough to cover all the cases of motivational indifference: there are cases in which people, even not being motivated to act in accordance with their moral judgements, can be considered rational or even *entirely* rational. Therefore, it is said that rationality[61] is not the right condition to secure the necessary relation between a moral judgement and respective moral motivation, therefore, the RI fails, motivational internalism is false.

The problem with this strategy is that these criticisms are based on a different conception of rationality than that of RI. So this approach only shows an apparent thing: RI does not work with the conceptions of rationality more or less randomly chosen by its critics. It is not to say that none of the critics tried to approach the RI with the conception of rationality, and, accordingly, irrationality, that is supposedly presumed by the RI. The various philosophers have forwarded this criticism from different perspectives on what rationality amounts to: the "follow-through" account, the instrumental and the (supposedly) common-sense conceptions of rationality. However, they all share one crucial feature: they attribute to rationality, as its core element, the normal mental functioning, whereas irrationality, for them, necessarily indicates abnormal or impaired mental functioning. It is true that in many cases the internalists and externalists list the various cases of mental malfunctioning as the apparent cases that internalism must account for if it is to be held plausible, but it is not sufficient to conclude that irrationality has to be

[61] Or at least some plausible conception of rationality. Any other definitions that could account for all the cases of motivational indifference, it is argued, are either *ad hoc* or issue in other serious problems (e.g., Zangwill also argues for the latter in 2008: 116). So the same conclusion follows anyway.

identified with mental abnormality. I will argue that this element is not a necessary part of "irrationality" at all – at least the way the RI understands it.

The supposed equivalence of irrationality and mental abnormality. Let us see the described criticism at work. E.g. Elinor Mason supposes that according to internalism, "it is abnormal in some way not to *do* the action you believe you ought to do" (Mason 2008: 150). However, she argues that we can imagine the whole scale of indifference, at one end of which we find people with brain damage, some more familiar cases of accidie, rage, grief and laziness in the middle, and the wilful ignoring at the other end. In other words, it ranges from what "normal agents wouldn't do", "cases of faultiness", the "abnormal" (which she considers to be what the RI can in some way account for), to the weak-will or wilful wrongdoing which is "perfectly normal, and depressingly common" (Mason 2008: 150-1).

Actually Mason's conception of rationality is quite close to the one employed by the RI. Mason distinguishes between *theoretical*, *means-end* and *follow through* conceptions of rationality, where the latter is "a matter of believing what you believe that you have reason to believe, or doing what you believe you have reason to do—i.e., following through" (Mason 2008: 147). However, the very classification she introduces and formulation of the *follow through* principle is enough to indicate that she understands it in a different way than the RI presupposes. Also, she still makes this presupposition about the essence of (ir)rationality, which lets to align her with the rest of the critics. Her main point here is that the cases of indifference do not necessarily indicate cases of mental impairment (and, in addition, even the cases of impairment might not be what we would call cases of irrationality), and irrationality is being identified by the RI with exactly just that. Therefore, she makes a conclusion that as "internalists have given us no reason for thinking that not doing what you think you ought to indicates a problem with the agent, so no reason for believing in motivational internalism" (Mason 2008: 153).

Zangwill and Strandberg also claim that the cases of moral motivational indifference they present do not seem to be cases of irrationality. Zangwill agrees that the listless or the depressed are obviously irrational, but not some others who are just "morally cold", "bad" or otherwise "rationally indifferent". The latter seem to be "perfectly content and well balanced", "even … quite happy", "normal", their "mental faculties … seem to be in order" (Zangwill 2008: 113-114).

According to Strandberg, the "term 'irrational' is used to categorize various failures of mental functioning", but "the examples … [which he has discussed] all provide evidence that it is not conceptually necessary for the person in those examples to be mentally malfunctioning in any relevant way" (Strandberg 2012b: 35).

Rationality as coherence. But what about the conception of rationality that the RI implies, what does it amount to? I claim that rationality for the RI is and should be identified *primarily* with *psychological* coherence. All the requirements of rationality can finally be reduced to requirements of coherence[62]. It is not a new idea, but perhaps not taken seriously enough. The various authors, for a clear example, John Broome in (2010), Donald Davidson in (2004a), Smith (1994, 1996, 2001, 2004b, and elsewhere), when talking about (ir)rationality talk about the inner (in)coherence or (in)consistency of mind. On this view, rationality is just taken to be a notion defining the relation between some person's psychological states in terms of coherence. And so the different conditions for rationality can all be spelled out in terms of coherence of the various kinds. If so, this would mean that there can be principles of rationality for connecting different kinds of states or sets of states of human psychology by the same type of relation (coherence). Then, even practical and theoretical rationality would not be differing substantially, the difference in

[62] An important supplement/reservation must be added: coherence is necessary for rationality, but another element is usually needed. We will supplement this account further in the text.

labels would only signal that coherence is required between different kinds of psychological states, or elements (say, beliefs of different kinds, or beliefs and desires, etc.). With respect to what elements should cohere or how[63] for a person to be identified as rational, we could analytically discern different kinds of rationality requirements, or principles.

If rationality is understood this way, then the concept of *full* rationality in the practical context boils down to the pervasive requirement of coherence among all of the relevant psychological elements involved in a certain practical decision. That way, the idea lying behind the RI is rather simple, and there is nothing contentious in attributing irrationality to the addicts, other indifferent individuals of the kind and even to people without diagnoses: those, breaking requirements of rationality, are simply incoherent (and not necessarily mentally malfunctioning, abnormal).

The possibility of such a conception of rationality, however, should have roots in our everyday language usage, as not only motivational internalists themselves, but also some of their critics turn to this court of appeal for the evaluation of plausibility of the RI claims. And I suggest that such an analysis of rationality, as roughly sketched above, is available.

About the method. Remaining faithful to the common-sense understanding of our most essential concepts, we have to hold that the analysis of "rationality" should be such as not to contradict the ordinary usage of it. It seems that when we want to define some concept (which is precisely what we are doing here) we are looking for what is common to all of those instances that we apply this concept to. Our linguistic intuition helps at least to indicate the cases (of positive or negative importance) worthy of attention for that analysis. (However, the final result of such an analysis can be such that ordinary language users need not acknowledge that they mean that or only that by the concept in question.) This is the part where folk intuitions and instances

[63] For example, interpersonal coherence (of beliefs or preferences), intrapersonal coherence (of preferences, etc.), intrapersonal synchronic or diachronic coherence etc.

of the usage of the concept are important. The rest, though, is philosopher's work. The task of a philosopher is to purify the concept, removing from it the inessential elements of meaning and the contradictions that sometimes contaminate folk conceptions. Again as Smith claims, "philosophers' theories do not generate answers that are different in kind to the answers ordinary folk give to moral questions. They are *merely* more technical and more systematic" (Smith 1994: 2).

One could say that the result of such a philosophical analysis that begins with a folk concept is a technical term (product of – or of use to – those that are skilled in clarifying concepts). But "technical term" can have some negative connotation: it is supposedly an artificial term that ordinary people do not use. However, "technical" should only be understood to mean a purified version of the corresponding folk concept, where the latter implies that residues of meaning or unessential connotations are present along with the meaning.

So the procedure of finding definition of "rationality" should go along those lines as well. The linguistic intuition of the competent language users can enable us to circumscribe the range of instances that should be taken into account, i.e. instances of both the "rational" and the "irrational". Then, the philosophical purification of the concept begins until we arrive to the definition.

Analysis of rationality. It seems first of all, though, that the word "rationality" is itself a philosopher's term of art, not so much a word used by the folk. For example, in his work Gert claims: "Of course I do not mean to appeal to intuitions about the use of the very word 'irrational', much less to the phrase 'subjectively irrational'. The first of these is rarely used by normal people, and the second is a technical term" and "That is, 'subjectively irrational' is meant to collect the spectrum of actions that range from 'silly' and 'stupid,' through 'boneheaded' and 'a bad idea,' all the way up to 'crazy,' 'insane,' and worse" (Gert 2004: 143). I can only agree with that, and, taking over Gert's idea, rather look in the everyday language for the words either

expressive of the same idea as "rational" or at least partly expressive of it. I shall look for the words, which are expressive of success or failure to adhere to some kind of requirements of reason.

In everyday language "rational" may correspond to "prudent", "wise", "clever", "sound", "sensible", "reasonable", "sane", and the like. In other words, we may categorise actions or agents with these attributes as "rational". Of course, each of these words has wider meaning than "rational", as well as differing connotations (functional, emotional or other kind of nuances). They might even have more than one meaning; however, roughly, we can think like this. It seems that "sensible" is that which judges or acts in accordance with relevance to the situation as represented to one by one's senses. "Clever" may be that which manages to find the relevant means to some end. "Prudent" is probably the one who presently acts so as not to compromise one's future interests. And so on. From this, it seems not too far-fetched to notice that they all share part of their meaning or, at least have a family resemblance: they all signal an instance of coherence in between some elements or sets of elements (decisions-senses, means-ends, present interests-future interests, etc.).

As for "irrational", there are several words partly corresponding to it in everyday language, primarily, "silly", "stupid", "crazy", "insane", "nonsense", etc. According to the Merriam-Webster dictionary, we can find such definitions or parts of them: "exhibiting a lack of common sense or sound judgement", "contrary to good sense", and so on. These irrationalities are due to the discrepancies with respect to the standards or to those, who hold to/embody those standards; actions fail to cohere with the standards (of reason). In other words, the aforementioned concepts are used to signal situations where one of the requirements of coherence is infringed, i.e. when there is some kind of incoherence in between some elements or sets of elements within a person's mind[64]. Once again, the meanings of these words

[64] "A person's mind" does not necessarily mean that it is an actual person's mind or that the incoherence is apparent from the first personal point of view. Standards may

are not equivalent to that of "irrational", but wider. Also, we can notice here that rationality of some action or agent can be judged against some intersubjective standards, not just against the knowledge of that particular person's current goals (and this is to the contrary as to what the adherents to the narrow view – instrumental rationality – could agree with)[65].

So "irrational" neither explains the error nor is used to evaluate the mental status or character of a person that it is attributed to. "Irrational" just records an error and categorises it: the one of incoherence. Presumably, irrationality can explain why the necessary relation between the moral judgement and motivation does not hold (the agent is not rational), but irrationality itself must be explained – by naming its causes or otherwise.

Certainly, mental malfunctioning can be such an explanation, but it is neither a necessary, nor a sufficient condition for irrationality, as we know that addicted people can quit their addictions and the mentally normal be weak-willed or that addicts even under the influence of their addictions do some rational actions, as well as the depressed are not entirely irrational. Smith agrees with Stocker: "The point is not that agents suffering from such maladies are necessarily irrational: they may or may not be" (Smith 1994: 155)[66].

As far as I am concerned, the various terms for mental conditions categorise a *recurrent* behavioural pattern (whereas "irrational" is *primarily* used for singular actions). To call somebody "depressed" or "addicted" is to

---

often be thought of in terms of an idealised person's mind; a gap may be, e.g., between the latter and the actual person's mind.

[65] But it is important to show that the roots of such a different than the narrow means-end rationality understanding are to be found in common-sense understanding. On my view, the means-end rationality conception as the sole legitimate conception of rationality is a weird dogma that needs to be rejected.

[66] It is rather that "Desires are irrational to the extent that they are *wholly and solely* the product of psychological compulsions, physical addictions, emotional disturbances and the like; to the extent that they wouldn't be had by someone in a non-depressed, non-addictive, non-emotionally disturbed state" (Smith 1994: 155). This means that it is only those desires that cannot possibly be shared by the well mentally functioning and the impaired, are necessarily irrational. The ones that can be shared can be rational or irrational – it depends on other things. In other words, irrational desires are those that are had by, e.g., the depressed *as* depressed, the ones on the basis of which they are characterised as depressed.

categorise a recurring psychological pattern on the basis of the character of their recurrent errors of incoherence (between that person's own best practical judgment of some kind and motivation to act accordingly). The depressed lack the relevant desire or perhaps a desire of a significant strength for self-regarding actions[67] and in the addicted the desire for drug (or for a certain state of psyche that certain drugs enable) is prevailing. But these or other similar labels do not deem these people for *complete irrationality*.

To call somebody "irrational" is primarily to record somebody's singular action[68] as falling short of one of the requirements of rationality (coherence requirement of some kind). So far, I have not discovered, therefore, that "(ab)normality of mental functioning" should be a necessary part of the meaning of the "(ir)rational"[69].

We can even go further and look at our own everyday lives. How many times per day, being mentally well-functioning, we act irrationally? Perhaps when we procrastinate to do something because of a fear to fail or while trying to avoid some, even minor, inconvenience? Or maybe when we are being lazy or just tired and so do not pick up urgent tasks that we acknowledge it would be best for us to do now? Or when in the morning the alarm clock goes off and you turn it off telling yourself that you will be up in five minutes, at the same time not believing this at all; perhaps even knowing that it will not happen, and knowing it is best for you to get up right now, but not doing it. When you knowingly succumb to the lure of advertising and buy something you do not

[67] Cholbi in his paper (2011) claims that empirical evidence points to the conclusion that the depressed usually lack in self-regarding motivation rather than the moral one, and that is to the contrary of what is popularly presupposed in the moral internalism/externalism debate.

[68] I sometimes say that "(ir)rationality" can be attributed to actions or agents, where there is not much difference between the agent and action: you are what you do. However, I advocate the view that one action is not enough to define an identity, therefore, "(ir)rational" first of all describes agent in face of one's singular action, and does not give an overall evaluation of one's character, unless in the context where the agent is evaluated in relation to one's more recurrent actions or patterns of actions. E.g. in popular usage: "He is terribly irrational: always does what he feels like at that moment".

[69] Even in cases of the aforementioned "insane", "crazy", "nonsense" and the like.

need? Are you being irrational? I would say that in all of these and many more cases we are irrational, and even by our own lights, if we are sincere enough to acknowledge it.

However, it seems that most of the time we are rational rather than not, because the irrational practices are not the norm, they need justification, explanation, we label them; and the world functions well enough instead of just falling apart in chaos as we stick to rules, promises, standards, commitments, etc. E.g. *most of the days* we get up and get to work on time and we resist lots of immediate temptations (such as to have a nice after-lunch nap, to go out and just to have fun, to eat cake, to skip lectures, to snap at our superiors, not to do chores, not to visit the annoying relative in need, etc.). Teachers teach, planes fly, parents take care of their teenage children, surgeons operate on people for hours, surprisingly many people survive the daily participation in busy traffic, and so on – despite the many impulses, desires, passionate emotions, naughty thoughts, tiredness, low mood or unwillingness to do that which we judge to be the right thing to do.

Therefore, to the contrary of the position of the RI critics, I hold irrationality to be a pervasive, but not an overwhelming phenomenon of everyone's daily life, not just some abnormality that necessarily happens only to the psychologically damaged or malfunctioning. I claim that the group of words (necessarily or not) referring to (or possible to categorise as) "irrational" is wider than that which would refer to the "abnormally mentally functioning" and that the latter class only partly intersects with the former. What matters to the attribution of "(ir)rational" is whether the coherence relation of some kind holds (or not), not whether the person is functioning normally. To put it otherwise, *it is not in virtue of the poor functioning of the brain that one is irrational, but in virtue of one's psychological states being incoherent* (even if the first may cause the latter). And the poor mental functioning is among those factors that sometimes can explain – not necessarily as causal explanations – the incoherence. As Korsgaard puts it: "Rage, passion, depression, distraction,

grief, physical or mental illness: all these things could cause us to act irrationally" (Korsgaard 1986: 13).

So at this point we are already able to answer the critics of the RI that their criticism based on the supposition that rationality necessarily implies normal mental functioning, fails. It is quite on the contrary to what they claim: the ordinary usage of the "rational" and its cognates indicates that rationality as well as irrationality are attributes that pertain to the ones that function normally and are "even happy" equally well as to those who do not – depending on the characteristics of their singular actions or decisions. Irrationality is not a diagnosis, it is because of the diagnosis that it can be pardoned, in some sense justified or at least understood.

3.2. Kinds of rationality and their relations

Criticism towards claim (2). Similarly, but even more pressingly, goes Strandberg's argument targeted at claim (2). He suggests we consider some cases in which a person has more than one normative reason for action. In one such case, a seriously ill person is presented with a certain available medical treatment and its side effect. Then, she has two incompatible reasons: to ϕ (act being "to accept the medical treatment", for a reason that it will save her life) or to ψ ("to decline the treatment", reason being that because of the treatment she will not be able to drink coffee for one minute). According to the normative internalist claim, even if this person considers the reason for ϕ-ing to be "absolutely the strongest reason", and the one for ψ-ing an "*extremely* much weaker" reason, she has to be motivated to do both, "*in order to be entirely rational*; ... she must be *irrational* to a certain extent *unless* she is motivated to act in that way" (Strandberg 2012b: 33).

However, Strandberg thinks we can hold her *entirely rational* even if she is not motivated (even to some extent) to decline the treatment, or, on the other hand, that she might be so motivated, *even if* she is rational. So the consideration of the presented case shows that competent language users *need*

not agree that someone, not motivated to act on an extremely much weaker reason (that is, not motivated *to some extent*), is necessarily irrational. Therefore, the intuitive conception of (ir)rationality, to which, according to Strandberg, rationalist internalists purportedly appeal, cannot secure the conceptually necessary relation between all and every reason and motivation to act accordingly.

In so far as Strandberg relies on the conception of rationality that we already showed is misconstrued (i.e. rationality as normal mental functioning), we have answered his worry. However, there is more to this criticism: it seems that we can hold the person from the aforementioned case rational in the sense of coherent as well. Should the insignificant incoherence (not responding with motivation to the weakest of reasons) influence our judgement of the person as rational? This is a sensible question to ask.

Strandberg's own position is such that in the cases he considers we hold such a person *entirely* rational. I suppose that here Strandberg is criticising Smith's conception of *full* rationality, and does it appealing to our intuitive understanding of full rationality. So let us turn to Smith.

Smith: conceptions of full and practical rationality. Smith's "full rationality", though, is not and should not, as I will claim further on, be an intuitive notion, therefore, one cannot intuit whether somebody is fully rational of not. According to Smith, "the idea of someone's being fully rational is itself a *summary* notion. The role of this idea in the analysis is thus to capture, in summary style, a whole host of more specific platitudes about practical rationality" (Smith 1994: 155-156). The difference between full rationality and rationality of some other kind, say, practical rationality[70], must be highlighted.

Smith adopts a slightly reinterpreted version of the conception of *full* rationality given by Bernard Williams which is spelled out in three conditions:

(i) "the agent must have no false beliefs

[70] In its narrow sense, practical rationality consists in being motivated to do what one oneself judges to be right for oneself to do.

(ii) the agent must have all relevant true beliefs

(iii) the agent must deliberate correctly" (Smith 1994: 156).

Smith, though, explicates the third condition differently than Williams. As rational deliberation is taken to be a way of generating new and extinguishing old desires, it is to be such as to sanction only the desires of an appropriate kind. Smith believes that we deliberate, i.e. generate new and extinguish old desires "by trying to integrate the object of that desire into a more coherent and unified desiderative profile and evaluative outlook" (Smith 1994: 159). And this procedure is "straightforwardly analogous" to what Rawls says about beliefs. So Smith takes the third condition of correct deliberation to be the condition of attempt at systematic justification, in other words, he takes it to consist in a procedure very similar to the Rawlsian "reflective equilibrium": it is a process of systematic justification of our desires. That means that full rationality is defined by the idealised epistemic conditions (i – ii) and a requirement of coherence (condition iii explicated differently than by Williams).

First of all, these are conditions for reason and moral judgement formation, as for Smith the moral judgement consists in a belief that one would desire that one oneself ϕs in circumstances C if one had a maximally informed and coherent and unified desire set. So the conditions define, first of all, an idealised deliberator, not an actual deliberator. However, for a person to actually *be* fully rational, one has to, other things being equal[71], 1) *have* the desire to ϕ in C, 2) in the face of the aforementioned belief (that one would desire that one oneself ϕs in C if one had a maximally informed and coherent and unified desire set), and 3) that belief to be true.

So even if a person is motivated to do something that one believes one has a reason to do, but that belief is not true, Smith would say that such person's "overall psychological state cannot be maximally coherent" (Smith 1997: 100, n. 18), that she is not fully rational, but we would grant her

[71] Keeping presupposed that it is *because of* the belief, and not just accidentally, that the desire is had.

"practical rationality" narrowly conceived. Practical rationality requires us to have the desires that we believe we would have being fully rational (Smith 2007: 288). And this type of rationality is fully compatible with theoretical irrationality, "a failure in the way she forms her judgment as to what is desirable" (Pettit, Smith 1993: 59).

In other words, those who desire to do what they believe they have a reason to do, are *at least practically* rational, and if their beliefs are true (they would indeed desire precisely that, were they fully rational), then, other things being equal, they are even *fully* rational.

This analysis needs to be accompanied by a couple of cautions. "Fully rational" (as already noted about the "rational") does not characterise a person in general (as if one was immune to irrationalities at any point in time, or in all one's decisions, or rational "in general" or so), but only in relation to some action or decision. It means, one's certain action[72] is beyond rational criticism. Besides, "fully rational" here is first of all defined in relation to one reason, or in a *pro tanto* sense[73]. Thus a weird sounding result in Standberg's cases: one can be "fully rational" in relation to one reason, and not "fully rational" in relation to another. However, I believe that it is a minor linguistic problem, a price one has to pay for choosing as one's basic unit the *pro tanto* reasons. The final or overall "full rationality" of the decision or action (*all things considered*) would depend on the full rationality of each and every minor decision anyway.

As already said, we can talk about different "rationalities", or requirements of coherence between different elements or sets of elements of psyche (or elements *and* sets of elements). Thus the seemingly differing meanings of "rationality" (and, accordingly – of "irrationality"). One person

[72] It may very well be that one should explicate requirements for full rationality even more, that is, add that the strength of the normative practical judgement should cohere with the strength of the corresponding motivation to act accordingly. I presuppose this here as I argue for the need of it in this dissertation.

[73] "Action", therefore, is not a description of an actual action, but of the possible one – it is a normative description. We are discussing, for the moment, the normative aspect of it.

can judge someone as rational, and another can judge the same person irrational, but in different respects (for example, as the one in whom the means cohere with the goals had *vs.* as the one in whom the goals had do not cohere with the goals to be had by her own lights, etc.). However, the fully rational is the one who satisfies all of the relevant requirements of coherence and so is immune to any further rational criticism (in relation to a specific action). We can talk about "rationality", of course, as about "full rationality", having the (pervasive) requirement of coherence in mind. But equally well we can, analytically, talk about rationalities, where "rational" signals that *some* of the coherence requirements has or have been met, i.e. "rational" being used as a narrow notion indicating coherence of *some* psychological states of an agent. If we think about rationality in the wide sense (as "full rationality"), then we can even talk about *degrees* of rationality.

A note to Smith's account of rationality. As for a relation of Smith's account of rationality with an instrumental view of rationality, the first one is not limited to the second. A maximizing conception proposes that a rational course of action for an agent is that which maximally satisfies her desires, or what "it is rational for an agent to do is therefore relative to what she wants most to do" (Smith 1994: 130). In one of his more recent papers Smith says that requirements of instrumental rationality are not "all there is to practical rationality" (Smith 2004b: 109). Instrumental rationality, according to him, is best understood as "a requirement of coherence on an agent's non-instrumental desires and means-end beliefs" (Smith 2004b: 93) and as such is not sufficient to preclude rational criticism of the agent's choices. It is so because in the instrumental account of rationality the non-instrumental desires and means-end beliefs that the agent has, as well as the values of their quantitative characteristics, are taken for granted. The principle of instrumental rationality is *insufficient* to guarantee the *complete* rationality of choice, it can do this only contingently. Therefore, only more global requirements of coherence could rule rational criticism out completely.

So Smith's account of rationality, which is referred to as that of *full* rationality, is indeed openly more robust than the instrumental account. In face of the considerations presented by Smith with regard to the limitations of the instrumental account of rationality, it can seem a good idea to redefine the concept. However, there is one precaution against it which, according to Zangwill, should be taken into account. He claims: "Expand the concept of rationality beyond its instrumentalist core and it may well turn out to be irrational, in that sense, not to be moved by moral considerations. But by itself that achieves little beyond a terminological redescription, since someone might not care about rationality, so conceived" (Zangwill 2008: 116). The possibility of mere redescription of the problem means the turn from the question "why to act morally" to "why to act rationally". In that case the normativity of rationality should be justified (whereas the normativity of the instrumental rationality, we should perhaps understand, is evident?). In fact, it is true that RI has to justify the normativity of rationality. It is natural having in mind that the RI puts reason into the basis of morality, thus the shift of the question to a more basic level[74]. But it is not obvious that this is a disadvantage of the theories in question. We will turn to the question of normativity later on.

However, another thing that Zangwill adds can be of more guidance in selection of the better theory in the internalist/externalist debate. He claims that the two available explanations to the question of why people are (or fail to be) motivated are (a) the presence (or lack) of distinct moral desires; (b) adherence to some non-instrumental rational norms (Zangwill 2008: 118). And the right explanation should be that of the folk, as he believes we do not have reason to distrust it (to which internalists could, to a big extent, subscribe). Zangwill believes, though, that the folk explanation would be (a), because the indifferent people hold themselves to be rational, and nevertheless indifferent (Zangwill 2008: 119). Thus he concludes that even if such rationality (as advocated by

[74] It is obvious that finally something will have to be held "normative full stop", because indeed questions should end somewhere or we will be condemned for endless regress. The question just is where we should stop. The RI goes far: it puts normativity deeper than the instrumental rationality.

internalists) existed and we had the right faculties to adhere to its requirements, it would be irrelevant to the explanation of the phenomenon of moral motivation.

What one should agree with is the formulation of the problem and perhaps the most important criterion for choosing between the theories and concepts. But it is far from obvious that the answer of the folk would or should be (a). Especially if we think that "rational" for the folk does not necessarily mean "mentally normal" or "instrumentally rational" – to the contrary as to what Zangwill supposed in his arguments.

Another thing Zangwill appeals to is the belief-desire model of action explanation that supposedly is a folk explanation of actions and which supposedly is advocated by the motivational externalists. But here again I want to object: it is not true that internalists deny the belief-desire model, the difference rather lies in different treatment of the relation between the beliefs and desires. The rationalist internalists believe in the possibility of the necessary relation between the two types of mental states (in virtue of believing in the potential of the faculty of reason), whereas the externalists do not. So again, it is not apparent that the belief-desire model that the folk supports should be read in externalist (so-called Humean) manner.

Less than full rationality and satisfied intuitions. If we can accept the analysis of full rationality and the accompanying considerations that I presented so far, then it should be clear that any single linguistic intuition of competent language users cannot reliably track "full rationality", and that "rational" is usually used to denote only one or another instance of (not full) coherence. I claim that an adequate notion of "full rationality" is to comprise *all* those instances of coherence that are traced by the competent language users in their usage of the corresponding folk notions. Rationality is defined positively by words expressive of coherence, and negatively – by words expressive of failures in coherence. However, none of the folk notions taken on their own can define and no separate intuitions can track *full rationality*.

Therefore, criticism based on the supposition that intuitively we hold people fully (or entirely) rational has no force.

Moreover, if rationality can be attributed to people exhibiting less than perfect rationality, then the rationalist internalists can share Strandberg's intuitions (that the person in the example is rational to choose the treatment), and still deny his conclusion (that rationality cannot secure the necessary relation). It seems that, as full rationality consists of a whole set of requirements of coherence, irrationality can occur as infringement of any one of these. Smith uses such phrases as "full rationality" and "pure practical rationality", "local" and "global" coherence, and "even more global requirements of coherence", setting even the "minimal standard of local coherence"[75], which indicates the existence of quite a spectrum of rationality. This means, that referring to the examples in Strandberg's paper, one could agree that we do find people rational if they are motivated to act in the way, backed by the "absolutely strongest reason". However, we may not hold such a person *entirely rational*, as being exempt from any rational criticism – if only because she is incoherent with respect to one's weaker reason. Though in fact, to hold one *entirely* rational we should know much more (we have clarified the conditions for full rationality before). However, to be sure, for the RI to be true it is enough that the person is *practically* rational (one is motivated in accordance with one's normative judgement), it is not necessary that she is *fully* rational.

So we have an appropriate answer to Strandberg's claim that "Even if the considerations I have offered do not defeat (2), they provide evidence against it, since they suggest that competent language users may reasonably doubt it" (Strandberg 2012b: 36). The competent language users will attribute rationality to the person in question, recognising one's coherence in one respect, but they can equally well attribute this same person irrationality in another respect, or say, all in all, that such a person is rational, but apparently not entirely; maybe – irrational to some extent. The "fully rational" self is exempt from rational

[75] Necessary for somebody to count as an agent at all (Smith 2004b: 107).

criticism; however, our less than fully rational selves can be vulnerable to rational criticism because of some or another infringement of coherence requirement, and still be rational as complying with some other coherence requirement(s). For example, people can be practically rational, that is, exhibit coherence of one's normative belief and desire to act accordingly, and "may still fall far short of full rationality: that is, their desires may not yet be maximally informed and coherent and unified" (Smith 1997: 100, n. 18).

What the folk really needs not to adhere to, is that such a person is somehow *globally*, *totally* or *very* irrational. But this can be accepted by the RI as well. However, competent language users do not need to *intuit* that any person is fully rational, for that, they would need to *reflect*.

3.3. Formal and substantive accounts of rationality and their implications

Criticism based on the *required/permissible* distinction. Now, the normative internalist thesis (2) on the first of its interpretations claims that "it is conceptually necessary that, for any action ϕ and rational person S, if S judges that she has a normative *pro tanto* reason to ϕ, then S is motivated to ϕ". Let us remember the aforementioned Strandberg's point: he points out that when a person has more than one normative reason, the normative internalist thesis requires that person to be (at least to some extent) motivated to act in all of those ways backed by the respective reasons, on pain of being irrational. He presents us with several cases, one of which I have already introduced and which is supposed to provide some evidence that claim (2) on the *pro tanto* reading does not hold.

This same reproach can be read in a slightly different manner than in the previous section. Strandberg claims that in all those cases that he has considered "it seems plausible to see it is *rationally permissible* for the person in question to be motivated to perform the action at issue, even while she need not be *rationally required* to be so motivated; neither must she be *irrational* to

the extent she fails to be so motivated" (Strandberg 2012b: 35). Meaning, that it seems wrong to say that it is rationally *required* that the ill person is *motivated* to decline the treatment (even if the required motivation was very weak).

First of all, this criticism advanced by Strandberg and other theorists he invokes, such as Ralph Wedgwood and Joshua Gert[76], relies on a plausible distinction between the rationally permissible and rationally required actions (accordingly – between purely justifying and requiring functions of reasons). Gert seems to accept the distinction between justifying and requiring functions of reasons as mirroring our common-sense perception of the differing degrees of pressure of the various reasons and of differing reactions of people to our succeeding or failing to respond to those different reasons. It makes sense to think that some reasons for actions merely justify actions, so that the respective actions are not rationally required, but only permitted. Then, if one does not perform a permissible action (even if there are no countervailing reasons), one is not guilty of irrationality. The other kind of reasons, though, not only justify, but also rationally require actions, and not acting in accordance with these reasons (in case there are no stronger countervailing reasons), does mean being irrational.

So far so good, but it is worth asking what enables the distinction to get its specific form, i.e. how can we know where to draw the line: what is permitted and what is required, which reasons are which. The theories in question, such as Gert's and Smith's, define reasons in terms of rationality, therefore it will be one or another account of rationality that allows drawing the line.

Distinction as enabled by substantive accounts of rationality. By "substantive" I mean such an account of rationality which specifies which contents of reasons are considered to be rational, for example, avoidance of

[76] Strandberg (ibid. p. 35, n. 17, 18) is referring to Wedgwood (2002: 349) and Gert (2004: 143).

pain. Whereas "formal" requirements of rationality or reasons, for that matter, do not list any specific reasons, but rather present a procedure for determining which particular considerations count as reasons. Here I borrow Parfit's terminology: "To be substantively rational, we must care about certain things, such as our own well-being" *vs.* "To be procedurally rational, we must deliberate in certain ways, but we are not required to have any particular desires and aims" (Broom and Parfit 1997: 101).

It seems that logically, the one and only means to get a three-modality structure (permissible-required-forbidden) is to introduce requirements of a substantive kind. Then, that which satisfies the requirements, is required (e.g. it is rationally required to avoid pain), what mismatches the requirements, becomes forbidden (it is rationally forbidden to seek pain), and the rest (in our case it is actions), which neither fit, nor infringe (or which exceed) the specified requirements, fall under the category of the permissible (it is rationally permissible to do that which does not cause pain for the agent). That kind of principle, that is, a substantive principle of rationality, ranges over a rather small number of actions, making the group of forbidden ones equally small, and the allowed ones consist in whatever number is left unclassified by these two categories.

This way to enable the distinction is chosen by Joshua Gert (as well as is advocated by his father Bernard Gert). According to Joshua Gert, such accounts of rationality "offer a list of reason-providing considerations, such as pain, premature death, knowledge, ability, and so on" (Gert 2004: 164). He is well aware that philosophers more often frown upon substantive accounts like this and prefer to specify what is rational by introducing formal constraints; however, he considers such unfavourable attitude to be just "a sort of prejudice". To be fair, though, I have to mention that his own account set forth in the monograph *Brute Rationality: Normativity and Human Action* (2004) does not propose any such list of the aforementioned kind, but "[account of objective rationality] only results in the *extension being specifiable* in terms of a list" (ibid). I hold that this "extension being specifiable in terms of a list",

however, does mean that Gert's account *is* substantive, because namely this "extension" enables him to assign some reasons, expressive of the respective requirements, the requiring function. And it is rejection of this kind of reasons that indicates agent's irrationality. The rest of reasons (the justifying ones) make actions intelligible, but irresponsiveness to them does not indicate irrationality (even regardless of there being only one reason for action at all).

Of course, as already mentioned, according to this view, *motivation* is respectively either permissible or required – *depending on the action* (whether the action is required or permitted). That is why from this point of view, being faithful to Smith's *pro tanto* reading of reasons and sticking to Strandberg's example about accepting or declining the vital treatment, acting according to the second of the reasons (not being able to drink coffee for one minute) is only a rationally *permissible*, but not a *required* option. And it is not required both from Gert's perspective (there is no requirement to drink coffee, whatever value it represents), and from the common-sense point of view (this is Strandberg's approach). Whereas avoiding premature death *is rationally required* from Gert's point of view, even if common-sense response here can differ, but that will not be relevant here. (What seems more surprising is that such action is not required even from the perspective of Smith himself: "Agents … are not rationally required to act in the one way or in the other" (Smith 2002a: 121).)

Therefore, from this theoretical perspective it is unclear why not being *motivated* to do what you are only rationally *permitted* to do, must make you irrational. "Irrationality" should apply only to cases of failure to respond to the rational *requirements*, but, according to Gert[77], Humean *pro tanto* reasons (and Smith is considered to be a Humean in this sense) do not have the function of requiring, but only that of justifying. Certainly, then one has to accept that none of the reasons – may it be the strongest or the weakest – as far as only rationally permitting actions, has to be necessarily embraced on pain of irrationality. Even not acting at all is a perfectly rational option.

[77] E.g. Gert in (2004: 175) and elsewhere.

This criticism *prima facie* applies. However, Smith's account of rationality consists, unlike Gert's, in purely formal requirements of reason. Then, it seems that logically, if the requirements are purely formal, then the tripartite modality (required-permissible-forbidden) structure is unavailable, leaving only the "required" and "forbidden" as two classificatory options. It is so because a formal definition will issue in a large number of rational options – so large that it will be practically pointless to make a list of them (of course, the number of options depends on a definition, but here I obviously have in mind the most common definitions, and not such which would narrow down the options to a number that can be counted on one person's fingers).

Think of a substantive definition of good and about a multiply realizable goodness defined purely formally. In the first case we will easily know what falls into the "grey zone" of the permissible, because we know which things exactly are good and bad. But in the second case it will be impossible to define the zone of the good exhaustively, and that which will not fit the definition, will be bad.

In that case, the number of the required options is bigger than while having substantive requirements, accordingly, the number of the forbidden options increases (it is most probably even much bigger than the number of the required options). But then, why do we find the category of "permissible" and supposedly (following the logic of Gert) no category of "required" in Smith's account? And how could possibly the "irrational" come into his picture?

Distinction as enabled by formal accounts of rationality. The answer is that there are other means that allow the distinction of the permissible and the required than that of introducing a substantive theory of rationality. If what is permitted is that which is not regulated, or required, by rationality, then Smith can well have all three categories.

Rationality in Smith's theory is to be defined as coherence, or, analytically, if we specify the requirement in accordance to which elements it applies to, then we can talk of rationality as a whole set of requirements of

coherence, and irrationality can occur as infringement of any one of these. What is required is the coherence of deliberator's psychology. In Smith's own words, "What is important to determining the rationality of agent's actions is rather whether they act as they believe that they have reason to, and whether these beliefs are in turn well-grounded" (Smith 2002a: 110). It means that rationality ranges over both the process of formation of reasons and the link between the reasons and motivation.

At this point we are most concerned with the latter and, accordingly, with what Smith (along with Pettit) calls the "failures of pure practical reason" (Pettit and Smith 1993). There can be several of these, but the requirement of pure practical rationality is that of coherence between the deliberator's "belief (true or false) that if he had a maximally informed and coherent and unified desire set he would want himself to do x in C" and "his desire to do x in C" (Smith 2001: 259). As already mentioned, not only coherence is a requirement on the content of the mental states, but also on their quantitative dimension – strength: "S's having a desire of a certain strength to do x in C, when he has the belief, true or false, that if he had a maximally informed and coherent and unified desire set then he would have a desire of that strength that he does x in C" (Smith 2001: 259-60, n. 2; with reference to Pettit and Smith 1993; Kennett and Smith 1994).

Therefore, the requirement of rationality is that of coherence that applies both to normative *pro tanto* reasons (is guiding their formation process) and to acquiring the relevant (by content and by strength) motivation to act accordingly. And this is a crucial thing to note: according to Smith, *rationality requires motivation* corresponding in content and strength to respective normative reason[78].

[78] Smith writes: "on pain of practical irrationality, someone who believes that there is a *pro tanto* normative reason to do x in C (a normative reason that may be outweighed by other normative reasons) must have some desire to do x in C (a desire that may be overridden by other desires)" (Smith 2001: 257). Again, here "some desire" does not mean "of whatever strength", but only that it does not have to be overriding with respect to other desires.

However, Smith's account accommodates the permissibility of actions and the purely justificatory function of normative reasons as well. That is so because requirement of rationality ranges on reasons and motivation, but not on the *content of values* of the agent, nor on the hierarchy of her preferences[79] (does not define what is to be desirable to any rational agent). Gert determines that rationality requires holding a certain part of the values superior, therefore, some reasons that embody those values, are superior to others and so they, accordingly, rationally require their embodiment in action. But for Smith, the content of the hierarchy of the desirable is not (or at least not obviously, or not *a priori*) regulated by rationality. So *rationally, every option, backed by a reason, is permitted; every reason justifies some particular action, but does not require it*[80]. Taking the example of the choice of the severely ill person, it is permissible for her to decline the treatment, because it would be justified by a reason, and it is permissible to accept it, based on another reason.

I understand that the interpretation I have given here can be taken by some to be weird, especially in the light of the definition of "reason" as that which is "required by rationality" (or by reason as ability). It may seem that there is some kind of confusion, because obviously that which is only permitted by reason, or rationality, cannot be required. But there actually is no confusion: what is a (normative) reason, is a reason in virtue of its form, not

[79] This is not to say that requirement of rationality cannot and does not rule out some of the values by purely formal restrictions (say, by some kind of interpersonal coherence requirement, where agents are defined by quite specified circumstances they find themselves in). However, it is quite different than specifying a list of values that all rational agents are to have or their hierarchy.

[80] Though the context is slightly different, I take these words of Smith to provide support to my analysis. Here he is talking about elements of Bernard Gert's account, of which he approves: "Agents have a free choice to decide in which way they will act, at least within certain limits. They are not rationally required to act in the one way or in the other" (Smith 2002a: 121). Of course, "within certain limits" is compatible with both readings: that the restrictions are coming from either substantive and formal accounts of rationality.
I should also stress that it is not certain that Smith himself would approve of my analysis in detail, however, having in mind his main arguments and the criticisms of his account, I hold this interpretation of mine to be valid at least as one of the possible readings of his account, as one of the logically possible ways he can take.

(primarily) because of its content. The rational argument about what we would want ourselves to desire to do under certain specified circumstances if we were fully rational, gives the answer a status of a reason. Because it is not obvious *a priori* what the content of those reasons would be. But every desire to do a certain thing that can be the content of a reason is permissible. The procedure, the universalizability of the desire grant the status of a reason. That which is a reason, is necessarily conforming to the requirements of rationality, but what is the *action*, or the content of motivation, is not *directly and a priori* required by rationality, it is *permitted* by rationality in virtue of being the content of some normative reason.

Thus, one needs to distinguish what (the faculty of) reason does in two different cases. First, under the circumstances of idealised rationality, when determining our normative reasons, it requires us to be coherent epistemically and with respect to our various desires (the three conditions of full rationality that Smith has set out). Here, reason does not require you to decide to wish to have a particular desire, say a desire to live, but it requires that you decide to wish to have a particular desire, in relation to the true information relevant for the decision and in relation to other desires of yours. Second, under circumstances of imperfect rationality, when forming our intentions in accordance with our practical judgements, reason requires us to be coherent and so to conform our actual desires with the hypothetical desires that we believe we would have were we fully rational, it *requires* to be *motivated* accordingly – whatever our reason.

So in Smith's case we can see that *motivation* is *required*, and not *also rationally permitted*, even if the respective *action* is. *Motivation* in response to a normative reason is always rationally required, it does not, so to speak, "follow" the *action's permissibility*, as in substantive theories. Rationality *justifies, or permits,* action through reason, but *requires* motivation. Therefore, not being motivated, at least to some degree (i.e. not having a desire of respective strength), to act in a way that one is only permitted to act, does

mean being irrational. And so, to the contrary of the critics, which presume that motivation always follows the action in its being permissible or required.

Is anything positive gained from this change of the status of what is required and what is permitted? I think so. On Smith's proposed view, it is possible that we are wrong judging that something is permitted or required, however, it is always required that we are motivated in accordance with our current normative judgements. This requirement preserves internalism, but it does not let to *a priori* determine and impose any kind of values as rationally preferable to others from some supposedly objective point of view.

3.4. Conception of reasons and their relation to rationality in the RI

Rationality and reasons. In this debate the various theorists are trying to give an underlying psychological explanation, or model, of our linguistic practices: what kind of metaphysical relations as implied by our moral language would obtain (in our actual world) in virtue of what psychological processes or relations. For an internalist, it is a question of how the non-contingent relation between the moral judgements and motivation is possible. Therefore, we should keep in mind that we are dealing with two levels – that of language and the psychological one, that of the states of mind. The first one is cast in terms of reasons, judgements and motivation, the second, in terms of beliefs and desires. However, they are not totally separated in the texts, but one should keep in mind the difference of relations (conceptual *vs.* causal, etc.) and of the possibility of different psychological models that can explain the same linguistic relation.

The question that has to do with this debate is that of the explanation of action. Though definition of action is itself a big problem, it is not unusual to think that to explain an action is to give the underlying reason for it. The problem then is what reason counts – explanatory (or motivational) or justificatory (or normative), and how they are to be defined.

As mentioned earlier, the internalist/externalist debate is *primarily* a debate about the *normative* reasons and their relation to motivation, or their potential to motivate. However, the motivating reasons enter the picture as well. So we should explain what both of them are for the internalists and how they are seen to be related. As well as what their relation is to our central concept, i.e. to rationality. It is rather apparent that rationalists define reasons in relation to rationality, yet deny that rationality is to be defined as mere responsiveness to reasons. How can this be?

What goes further, will be explication of the internalist model of reasons, their relation to action explanation and rationality.

Normative and motivating reasons. The "moral problem" that Smith gives an attempt to solve in his influential monograph (1994), is a conflict among the three separately plausible propositions:

1. Moral judgements of the form 'It is right that I ϕ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.
2. If someone judges that it is right that she ϕs then, *ceteris paribus*, she is motivated to ϕ.
3. An agent is motivated to act in a certain way just in case she has an appropriate desire and means-end belief, where belief and desire are, in Hume's terms, distinct existences. (Smith 1994: 126).

The first claim is essentially an analysis of moral judgement, the cognitivist claim, the second one is about the relation of judgement and motivation, the internalist claim, and the third one is about an analysis of motivation, a Humean view of motivation. Smith thinks we can solve the problem without rejecting any of them. He embraces the (3) Humean claim that motivation is produced by desires, but disapproves Hume in holding that no desires can be themselves produced by beliefs[81]. This lets Smith to connect

[81] Here I leave open a question of whether Smith can be held to be Humean in any sense at all. Smith says: "According to Christine Korsgaard for instance (1986), anti-

the second (through the first) and the third claims: moral judgements consist in beliefs that produce motivation, which, in its turn, consists of desires and the means-end beliefs. Supplied with a proper analysis, it means that the normative reasons and the motivational reasons both take part in the production of action – at least they both can (as there can be actions without normative reasons), and they may be connected.

For Smith moral judgements, as well as other kinds of normative judgements, *psychologically* are beliefs about normative reasons, beliefs that there are normative reasons to do something (that is their form). As Smith himself puts it, normative reasons are facts, requirements of rationality, "normative reasons are best thought of as truths: that is, propositions of the general form 'A's ϕ-ing is desirable or required'" (Smith 1994: 95), where the "desirable" means "I would desire to ϕ if I was fully rational". Therefore, moral judgements are beliefs that some course of action is desirable.

Motivating reasons, on the other hand, according to Smith, are better thought of as *psychological states*, not propositions or truths, so they are entities of a different *category* (in relation to normative reasons). Humeans think that motivating reasons are constituted by desires and means-end beliefs.

But as Smith says himself: "However a theory of motivating reasons is not the whole of a psychological theory, for, as we have seen, alongside motivating reasons there are also normative reasons" (Smith 1994: 130). How should we understand this distinction, if normative reasons also have their psychological realisation and are talked about in psychological terms as well?

Humeans assert, and Humeans deny, that reason can produce a motive. But, as we have seen, what is at issue is not this, but rather whether the reasons that produce motives are themselves relative, as the Humeans suppose, or non-relative, as the anti-Humeans suppose" (Smith 1994: 213). According to both of these criteria Smith qualifies as anti-Humean. And it is a question if Kantians, for example, are really against the idea of belief and desire being different entities and that desire is a necessary but insufficient element of motivation. That is, Humeanism as in claim (3) may not be a specific enough description of the Humean view. It is possible that it differs from Kantianism only on whether this motivation can constitute a description of action. But this is to be pursued elsewhere.

Is it really best to distinguish the two kinds of reasons in terms of different categories? The answer is to be found in the further considerations.

When talking of the normative and motivating reasons, a certain asymmetry holds. On the one hand, we can say that there are normative reasons for acting in some way or another, even if a person is not aware of them or on which one fails to act, on the other hand, we can explain one's action by citing the normative reason for which one performed the action. Therefore, normative reasons can always be understood as justifying considerations, but only sometimes they are also psychologically real, i.e. play a role in the explanation of the actual psychology of some agent when explaining what one did. However, it does not make sense to talk about there being motivating reasons of which an agent is not aware of as they are always psychologically real, they do not exist elsewhere than in an actual agent's mind. Accordingly, this asymmetry is represented in Smith's theory: one can act for a motivating reason without having any normative reason, but if one acts for[82] some normative reason, then one's action also has a motivating reason that can explain it[83]. So it is primarily this difference between the reasons that Smith chooses to describe as difference of category: normative reasons primarily and necessarily are truths, and motivating reasons are necessarily psychological states.

Smith thinks that "*motivating* reason" emphasises the explanatory dimension of the "reason" and downplays the justificatory, and "*normative* reason" – vice versa. Reasons make an action intelligible, but from different perspectives: normative – from the perspective of the normative system that generates that certain requirement (the *deliberative* perspective), motivating –

[82] Which is not to say "if there is a normative reason for one to do something in particular" (nor, in Smith's terms of (1994), "if one has a normative reason"). "There being" (or, accordingly "having") a reason may not be enough: "the mere existence of this normative reason is consistent with the claim that I am not in a state that is potentially explanatory of my behavior" (Smith 1994: 97).

[83] However, Smith would not say that motivating reason and normative reason in the latter case are the same. To the contrary of Broome who holds rather different views on reasons and action explanation in general (2009).

from the psychological state perspective in which the agent is (the *intentional* perspective) (Smith 1994: 96). Again, normative reasons primarily justify and only sometimes, when acted for, explain, and motivating reasons explain actions.

It is exactly here that a certain controversy springs even among those who restrict actions to rational actions. The controversy revolves around the question whether to cite a motivating reason, or intention (as some term it instead[84]), is enough to describe an action, i.e. if such intelligibility is sufficient and whether such an explanation enables to see an agent in one's role as a "rational animal", to use Davidson's term.

Davidson assumes that agent can be seen as rational only if she has done something rationally *justifiable*, but Smith thinks there are two ways to see her as rational, and it is *enough* to see her as in pursuit of a goal (therefore, to cite a motivational reason which is a teleological explanation of action). Smith thinks so because rationality is involved in forming both types of reasons, and people execute their rationality in relation to any of them. People can be more or less rational: being instrumentally rational is enough to be counted as rational, but it is not all there is to rationality as the goals, the means for which you are able to choose, can be irrational. So the goals, or the desire-component of the motivation, can have a "rational etiology", through which agents exhibit more rationality than otherwise. But "the possibility of giving such a rational explanation of desire … is in no way essential to an action's being an action. What makes an action an action is the fact that it can be teleologically explained by a desire for an end and a belief about means, something we can establish to be so under a conspiracy of silence about the rational etiology of that desire and belief" (Smith 2007: 296).

In Broome's model we have "motivation" replaced by "intention" and "moral judgement" with "ought-belief", but the similarity is obvious.

[84] I am aware that "intention" may not be identical to "motivating reason", especially as Broome thinks it is different from other kinds of motivation in being a stronger commitment to do what is intended.

According to him, act is explained by intention, and intention by ought-belief (which is an all things considered judgement, based on normative reasons). Therefore, the primary explanation of action is intention, which can be based on something or not. Davidson, then, faces a problem of akratic action (he takes it up in 2001): how is a weak-willed action possible, if its normative reason is not carried out? Smith criticises Davidson and manages to avoid his problem of akratic action: "He [Davidson] assumes, wrongly, that seeing an agent in this role [of a Rational Animal] requires that we see her as having done something that is rationally justifiable, at least by her own lights, whereas it requires, at most, that we see her as in pursuit of a goal that she has" (Smith 1994: 140-141).

For Davidson, an incontinent action approaches the sphere of unintentional behaviour and cases of split brain, such an agent is divided. Davidson speaks of subdivisions of mind, and "The breakdown of reason relations defines the boundary of a subdivision" (Davidson 2004b: 185). This understanding is close to Korsgaardian understanding of action as something for which an agent has to be united. She thinks that hypothetical imperative cannot stand alone (I will develop her views further on in more detail), so instrumental rationality is not enough for an action. The action is a *defective* rational action, a failure, an action that is bad *as* action[85] (Korsgaard 2009: 161). The more demanding understanding of "action" and "reason" enables thinkers of the latter kind not to call such actions and reasons for them proper actions and reasons. For example, Davidson says: "if the question is read, what is the agent's reason for doing *a* when he believes it would be better, all things considered, to do another thing, then the answer must be: for this, the agent has no reason" (Davidson 2001: 42).

One more similarity of Korsgaard and Davidson is the belief that an action is an intelligible object: "An action is an essentially intelligible object

[85] "The function of an action is to unify its agent, and so to render him the autonomous and efficacious author of his own movements. An unjust or unlawful action therefore fails to unify its agent, and so fails to render him the autonomous and efficacious author of what he does" (Korsgaard 2009: 161).

that embodies its reason, the way an utterance is an essentially intelligible object that embodies a thought" (Korsgaard 2008a: 228). So an action is an embodied reason – "described in a way that makes it intelligible" (ibid.: 227). And Davidson thinks that perception of someone as rational consists in viewing his movements as part of a rational pattern, however, in case of incontinence, "the attempt to read reason into behaviour is necessarily subject to a degree of frustration" (Davidson 2001: 42), and the cognitive dissonance is present in the agent: "the actor cannot understand himself: he recognizes, in his own intentional behaviour, something essentially surd" (ibid.).

But whereas such a more demanding conception of action has problems with explaining akratic actions and responsibility questions, it is not as terribly different from the less demanding conception. Both Smith and Korsgaard allow for degrees of rationality (thus, degrees or authenticity of agency), besides, one can say that one of them uses the term only in one's normative sense, whereas the other – also in descriptive (it is an action, even if it is not a perfect action that would be immune to rational criticism, it only satisfies the minimal requirements for an action).

As already mentioned, the metaphysical ties as implied by linguistic relations would obtain in our actual would in virtue of psychological relations, for example the intentional character of acting can only be recognised if the agent's action (bodily movements under a certain description) was suitably causally related to her attitudes that produced it. Action theorists discuss the problem of the so-called "wayward causal chains". A known example is of an actor who wants to play one's role, which is to shake as if extremely nervous. But once on the stage she becomes so overcome by stage fright that she cannot play her role, but only stands there shaking nervously. The cause of the action complying with her intention, however, was entirely fluky: her nerves caused just what she intended to do, thus we cannot conclude that she acted intentionally.

Therefore, even if the levels differ, the psychological one is apparently playing a substantial role in defining the other. Certainly, the nature of

relations on the two levels differs as well. And even if the premise of internalism is assumed, the underlying psychological states can stand in different causal or other kind of relations to each other.

As to the problem of wayward causal chains, Smith offers a certain suggestion which is not as interesting in itself for us[86], but which is important as amounting to the requirement "that the agent *has and exercises the capacity to be instrumentally rational in a very local domain*" (Smith 2009: 529). The requirements of such minimal capacity are the (minimal/necessary) requirements for agency (ibid). In other words, talking about non-flukiness of causal relations only makes sense if a person's substantive rationality (in the sense that people can be and are, at least sometimes, rational) is presupposed.

Relation of reasons and rationality. It is quite common in the various papers to find considerations about whether rationality for the rationalist internalists amounts to responsiveness to reasons[87]. Naturally, the considerations are followed by criticism that defining reasons in terms of rationality and rationality – in terms of responsiveness to reasons is circular or *ad hoc*. However, even if some authors, such as Parfit (Broome and Parfit 1997) and others would agree that rationality does amount to exactly that, both Smith and Broome answer in the negative: there is no deep, or necessary, relation in between the normative reasons and rationality. However, on the other hand, rationalist internalists do define reasons in relation to rationality. How can it be?

Rationality is not to be defined as responsiveness to reasons for several reasons. First of all, such a definition, as Gert points out, would indeed be

[86] I can just notice here that a possible solution to the problem of wayward causal chains is such as to answer the question of Setiya: "What is it about being-done-for-reasons – or being susceptible to the question 'why?' – that requires the presence of belief?" (*apud* Smith 2009: 525). It is exactly the means-end belief (its variations in the possible worlds) that enables to define the condition of non-flukiness, as a person is understood to act only if one is differentially sensitive to her slightly different desires and means-end beliefs.

[87] And here one can talk about both normative and motivating reasons – depending on the context.

circular in constructivist accounts: "if the relevant notion of rationality included a responsiveness to moral reasons, wouldn't such a contractualist account simply amount to complex but pointless circle? And how could such an account ever hope to shed light on the nature of moral reasons in the first place, since it would seem simply to presuppose them" (Gert: 2007: 172). It is especially true of substantive accounts of rationality, which lay out in advance what is the rational thing to do. Such circularity would make the definition of reasons into a comfortable one, excluding the possibility of the existence of reasons that are out of the range of the rational persons. Formalist accounts of rationality would be in serious trouble: not only would they not shed light on the nature of moral reasons, they would not shed light on rationality itself.

As Michael Smith notes, the nexus between *normative* reasons (for belief or action) and rationality is not deep, or in other words, not necessary. There can be normative reasons out of reach of the awareness of an actual rational person, as well as the subject can be rational without acting or believing for a good normative reason: one's behaviour can make sense because of good motivating reasons, but not because of one's acting on justifying, or normative, reasons. Broome (2009, 2010) and Coates (2011) emphasise the same asymmetry.

Smith distinguishes three ways of talking about (normative) reasons and three kinds of a person's relations to them. To say that *there are* reasons to believe or do something, is to say that there are "considerations that justify" believing or doing something, or "ways things are which make it rational for someone who believes that things are that way" to believe or do that something (Smith 2007: 283). The reasons that *a subject has* for believing or doing something are a subset of the first set, they are *available* to the subject. And if a reason is not just available to a subject, but also she believes or does something *for* this *reason*, that is, she is aware of it and it figures in an explanation of the right kind of her believing or doing it, such reasons are a subset of the previous subset (ibid.).

Reason is a fact, of which an agent may be aware or unaware, but rationality of agent's acts is determined not by whether an agent acts in accordance with a reason for action that there is, but by "whether she acts *in the belief* that she has an adequate reason for so acting, where that belief eludes rational criticism" (Smith 2002a: 114; emphasis mine). The reason that one believes one has may be false, as well as the evidence available to one may be misleading (but not by one's own lights!), thus, one can still be rational in forming a false reason or false judgement, and acting for those false reasons or on those false judgements while not acting for the (right) reasons which are facts. Thus acting for a reason and acting rationally may be just different things.

Now the confusion can arise because *in language* we shift between the first and the third personal perspectives as well as between the linguistic and psychological levels. Smith defines reasons in terms of *fact* rather than belief, so the ontological status of it is such. He insists on reasons being facts, even if in the internalist debate we usually talk about the believed reasons (and rationality). Hence the cognitivism: belief, expressive of a judgement that there is a reason to do something, can be true or false in virtue of its true or false reference to a fact.

Also, the relation that he claims to obtain between the motivation and judgement is that between a *belief that* there is a reason to do something (a belief expressive of a judgement) and the relevant desire. However, from the first person's perspective, as far as one holds one's belief to be true (even allowing the possibility that some newly available evidence could make one change one's own mind), one does not make a distinction between a fact and a believed fact[88]. If somebody asked why you did it, you would answer with pointing out the content of your belief and not by referring to your believing that content, or by pointing to your reason, not to your belief in a reason (again

[88] So to speak, if there is no reason for you to bring your own states of mind into reference, you do not do it, even if you acknowledge a possibility of such a future need. However if your belief is very uncertain or not supported by sufficient reasons, you can refer to belief – distancing your own mind from reality.

– unless you see now that back then you were wrong). For example, "because genocide is bad" instead of "because I believe that genocide is bad", "because killing innocent people is totally unjustifiable" instead of "because I believe that to kill the innocent is totally unjustifiable". To be sure, one could utter the latter versions, but then the morality we would be talking about would have the relative character which is not in sync with the categorical character of it as supposed by the rationalist internalists (and they suppose this based on the common-sense understanding of morality). It is perfectly clear that here we would be talking of opinions, not facts or truths.

Therefore, "belief" or "believed reason" do not figure in your language from the first personal point of view. It is only from the third personal view that we add reference to the states of mind. That is why Smith insists[89] on reasons being facts, even if they become relevant to the internalist debate when believed. It is like with the normative reasons: they are always truths, which only sometimes can be explicated in psychological terms. Reasons are primarily and essentially facts, and just sometimes they can be described on the psychological level as well: a reason is a reason, even if somebody does not know about it, but it is still a reason, when somebody gets to know it, when it is believed[90].

And from the third personal point of view we just do bring the references to states of mind in and give explanations in these terms, even on linguistic level: "he did it because he believed it was right". But "he had a reason to do this" shows recognition of the fact, whereas "he thought/believed he had a

[89] As in criticising Bernard Gert's account of reasons as "conscious rational beliefs" (Smith 2002a: 113-15). Smith claims to be here on the side of the common-sense conception that holds reasons to be facts.

[90] One should not think of reasons as something as objective as ideas in a platonic sense or similarly. What their factuality consists in will be laid out in the coming pages, as well as mentioned on other occasions in this text. But in short, they are constructed facts which get their intersubjective character from a procedure of idealisation: "Facts about what we have normative reason to do are constructed facts: they are facts about the desires we would all converge on if we were to come up with a maximally informed and coherent and unified set of desires" (Smith 1997: 97).

reason to do it" is a psychological description with an element of downplaying if not negating the status as a fact of that which is referred to by that belief.

Again, Smith's own formulations such as "reasons … are considerations which make it rational for him to desire to act in that way on condition that he believes that those considerations obtain" (Smith 2007: 289) emphasise the difference between the reasons as facts and their being available to the agent, and does not mean that the agent must believe that one believes that one has a reason. It is important to have this distinction right because it is easily confused and it will be important for what I say further on. From this citation we also see another thing. Even if it is rational to respond to reasons that one believes one has, or to respond to the conclusions from those considerations (from reasons), one is not necessarily rational responding to the reasons that there are (exist unbeknownst to the agent). Rationality is not an absolute notion, it is relational (not relativist!). It is applicable evaluating both ideal and defective knowledge and relevant motivation, but it always evaluates a human mind – be it perfect or imperfect.

If one asks what kind of facts reasons are, Smith says these are "facts whose status as reasons is conferred upon them by their relations to idealized psychological facts" (Smith 2002a). Let me cite: "the fact that an agent *A* can perform an action of a certain kind *K* in certain circumstances *C* by performing an act of kind *K** in those circumstances constitutes a reason for him to perform an act of kind *K** in *C* if and only if everyone, *A* included, would want that they themselves perform an act of kind *K** in those circumstances if they were fully rational" (Smith 2002a: 117). In other words, the normativity of those considerations that we call "reasons" comes from the fact that these are considerations of idealised kind, they lack the imperfections of people's everyday epistemic situations. The question could be why that which is privileged epistemically has this power, but let us put it off for the time being (it is explained elsewhere in this dissertation, Part II, Ch. 5). The fact that gets this normativity is the one about the desires of all fully rational creatures.

Turning to another author, Broome doubts that rationality is normative, that there are reasons to satisfy the requirements of rationality. However, he also embraces the view that rationality and responsiveness to reasons are different things. He thinks that we often have false beliefs about reasons, nevertheless it is rational to intend to do what you believe you have reasons to do, even if responding to reasons you would intend to do something else (Broome 2010: 288).

So we can see that rationality is not defined as responsiveness to reasons, even if we talk about the reasons that somebody believes one has. Rationality is best understood in terms of coherence, besides, one is required to be rational not just while responding to reasons one believes to have, but also in the process of their formation. As already said, reasons are defined in terms of rationality. However, rationality does not determine what the normative reasons are to be, as e.g. on some substantive conceptions of rationality (such reasons would then be avoiding pain and death, seeking one's own benefit or so). Conditions of full rationality just set the conditions for determining normative reasons and confer the normativity onto those factual considerations that thus become reasons. No circularity is involved though: in reason formation rationality operates on ideal agent's psychological states, and in acting – whether in theoretical or practical sphere – it operates on the actual states of agent's mind.

4. The possible implications of Moore's paradox to the RI

Moore's paradox. Another interesting instance worth of our attention are the statements that exemplify the so-called 'Moore's paradox', more specifically, one particular interpretation of it. A classic example of Moore's paradox (MP henceforth, standing also for the "Moore paradoxical" where adjective is in place) would be this (I here borrow all examples from Cholbi 2009, unless noticed otherwise):

(P) It's raining, but I don't believe it.

The two atomic propositions in the MP propositions are not straightforwardly contradictory, however, we find their conjunction puzzling. There can be several interpretations of the paradox, in the recent discussions, though, it is usually agreed that the paradoxicality is not due to violation of some conversational norms (Cholbi 2009: 496). MP statements are supposed by many thinkers: Kriegel, Heal, Shoemaker, Williams, Adler and Clark (as referred to by Cholbi 2009: 496) to inherit paradoxicality from their underlying mental states or attitudes. Therefore, in his paper of 2009, Cholbi follows Almeida (2001) and Adler (2002) in treating Moore's paradox as an instance of epistemic self-defeat.

The explanation that this position offers is such that a speaker, uttering an MP proposition of some kind, is "adopting contradictory attitudes toward the same body of rational considerations", "is epistemically at odds with herself" (Cholbi 2009: 500). The contradiction comes from a speaker's taking there to be sufficient evidence for asserting, say, that it is raining, but failing to believe what is warranted by exactly the same evidence: "The assertion of a proposition and the assertion of belief in a proposition are warranted by the same evidence such that in any case in which the assertion of a proposition is epistemically warranted so too is belief in said proposition (and, prima facie, the assertion thereof) warranted, and vice versa" (ibid). Thus the same evidence should suffice for both, and therefore, in case of MP propositions there is something "amiss in the speaker's state of mind".

In relation to the other available explanations of the MP, according to Cholbi, this interpretation "has the advantage of accounting for both the omissive version of Moore's paradox – the assertion of 'P but I don't believe that P' – and the commisive version – the assertion of 'P and I believe that not-P'" (Cholbi 2009: 500, n. 3) and he finds this interpretation the most independently plausible. Besides, such interpretation serves best to illuminate the paradoxicality of the moral equivalents (not in content, but in form) of the assertions in question (Cholbi 2009: 500, n. 2):

(Q) Hurting animals for fun is wrong, but I don't believe it.

(S) Hurting animals for fun is wrong, but I don't care.

Cholbi calls propositions like (S) moral Moore-paradoxical propositions and I will follow him in this. Certainly then, the moral MP propositions are to be held epistemically self-defeating as well. Cholbi himself does not specify if moral judgements are to be interpreted in the cognitivist or non-cognitivist fashion, however, I believe that the interpretation serves cognitivism especially well and fits into the parallel with (theoretical) MP propositions. And so in the context of this paradox I will treat the moral propositions in cognitivist fashion (cognitivism, after all, is one of the two focal positions of the present dissertation).

Assertions like (Q) can surely be seen as equally paradoxical as assertions like (P), but what about (S)? And what hinges on the plausibility of the analogy?

On the face of it, the possible counterarguments or at least doubts about the paradoxicality of (S) suggest themselves. For example, a cognitivist externalist could acknowledge the paradoxicality of (Q), because the mental state that underlies the first atomic proposition just is belief and the second atomic proposition is the belief of the contrary truth value, or, alternatively, the second one is the disbelief with the same content. However, statement (S) is different: care is not a matter of belief, there is no contradiction of the mental states or of their contents. Why should there be any necessary relation between a belief and desire, thus, why think (S) paradoxical?

Besides, the equivalent of (S) is

(R) It's raining, but I don't care.

Cholbi himself notes that assertion (R) is not paradoxical (Cholbi 2009: 495). Then why think that (S) is?

Let me offer a couple of answers to these possible worries. First, (R) surely could be paradoxical because of its context[91], however, it is true that it is

[91] But we are not interested in contexts ((P) may contain no paradox in certain contexts), we are interested in understanding the meaning and necessary implications of making an assertion, or judgement, of a certain kind.

not *as* readily paradoxical (or at least evoking some protest in some of us) without any specific context as

(S) Hurting animals for fun is wrong, but I don't care.

I may offer an explanation as to why it is puzzling, whereas (R) is not. Facts of the former kind (about the states of non-human affairs) surely influence our choices in the practical contexts, e.g., if we are planning a trip to the mountains or a picnic, then the fact that it is raining possibly will become significant. But one only should care about the non-human facts if one finds oneself under certain circumstances, i.e. if one has certain goals and these facts are relevant for reaching them. After we have answered the question of what to believe (whether it is raining), we can have further questions, for which the same answer will be relevant or irrelevant. The fact that it is raining may be for us utterly irrelevant and may not have any necessary implications. E.g. I sit home reading despite the weather or I want to have a walk regardless of the weather.

Moral judgements, however, are always judgements not about some non-human factual matters, but about human actions and practical contexts. The question of what to believe (what is true) is important, but it is always to be asked with the practical aim in mind, i.e. the answer to what is to be believed will be the same as to the question of what is to be done. The first is asked because of the need to answer the second, or, if you wish, it is the same question. It is so as far as one is always and inescapably in the circumstances of being a human, or having a goal of being human. Factual judgements do not imply any way of acting, whereas moral or otherwise practical judgements do[92]. Practical judgements just are behavioural directives.

Second, to reply to the aforementioned cognitivist externalist, I have to underline again that the puzzlement about (S) is not due to a contradiction of

[92] One may think that (S) is not paradoxical because the moral considerations are not overriding. E.g. I believe that hurting animals is wrong, but I care about something else (say about getting a safe cure that is tested on animals) more than this. However, as I said, we should look at these propositions without supposing any particular context, so in this case we do not know of other practical beliefs and so we do not take them into account.

the two mental states of the same kind or a contradiction in their contents. It is due to the incoherence in the attitude to the same body of evidence. According to Cholbi, in a looser sense, in which MP propositions can be held paradoxical, "paradoxes are offenses against reason, not against logic per se" (Cholbi 2009: 500). He joins Nicholas Rescher in claiming that paradoxes embody "dissonance of endorsements" and Saul Smilansky in defining them as expressing "the fundamentally alien relationship between the state of affairs and human reason". Other words that describe the paradox in the text are "lack of integrity", "inconsistency", "incoherence", "[lack of] authenticity", "unintelligible". In some places of the text Cholbi is even more explicit: "something amiss in the speaker's state of mind", "a rational misfire within the speaker's psychology, a misfire flowing from how the speaker conceives the relation between these two conjuncts", "irrational".

It means that paradoxicality arises as irrationality of the conjunction of the propositions, or, strictly speaking, rather as irrationality of people, because rationality is a characteristic of people and their actions, not of propositions. This and the characterisations cited enable us to see into the nature of rationality itself. The propositions themselves are neither contradictory, nor unintelligible. The confusion is due to the attitudes of the speaker, the paradox consists primarily in the inconsistency within the psychology of the speaker. The differences of rationality and other similar (epistemic) virtues become clearer. That, which is irrational, simply does not make sense[93]. In order to make sense, a further explanation is required. Irrationality does not necessarily indicate a flaw in logic and does not transgress the boundaries of that which can be understood by the human mind, however, it does indicate incoherence. Whether we can go further and accuse the irrational speaker of "hypocrisy", "madness" or something else, depends on further inquiries on our part and further replies (in words or actions) by the person in question, but the least we

[93] As in "I hear what you say, but it does not make sense" or in "I understand what you say, but it makes no sense" or in "I understand what you mean, but it does not make sense". I take it that rationality is a lot about "making sense".

can say is that one is saying something that does not make sense (again, without further explanations).

Thus Moore's paradox suggests that practical (moral including) belief and desire are necessarily related, but not directly, instead, related through normative reason. To be more exact, the paradox arises when belief and desire express contradictory attitudes towards the same body of evidence. Acknowledging that a certain body of evidence suffices to base one's judgement on means basing one's judgement on a good reason. The reason that is good for believing something to be a right course of action is also a good reason for desiring to do that right action. The one who finds there to be enough evidence to base one's practical judgement on, but not enough evidence to base one's intention on is incoherent. So again, the puzzlement arises from the differing attitudes to the same body of evidence. To bring out the incoherence, let us take the aforementioned assertions:

(Q) Hurting animals for fun is wrong, but I don't believe it.

(S) Hurting animals for fun is wrong, but I don't care.

Now, following Adler and Cholbi's interpretation, we can form equivalent assertions to these:

(Q*) Hurting animals for fun is wrong, but I don't have sufficient evidence for believing that hurting animals for fun is wrong.

(S*) Hurting animals for fun is wrong, but I don't have sufficient evidence against hurting animals for fun.

This nicely brings out the dual function of moral reasons in comparison to the non-practical reasons: the same evidence serves not only as a basis for belief, but also as a basis for desire. And it makes sense that one bases one's practical belief and desire on the same reason, as far as one is sincere. To put it figuratively, so far as one practises what one preaches.

The RI and the moral Moore's paradox. It should be underlined that the paradox holds only if we treat both atomic propositions of the assertion as first-personal, and not thinking that one of them is just a report. If the situation

is such that a person (*A*) reports what somebody else (*B*) endorses and (A) distances from it by the "I do not believe it" or "I do not care about it", the paradox collapses, as it does in the case where we have different reasons at play – one for belief, and a different one for desire.

One more feature of the MP is that its paradoxicality is content independent (Cholbi 2009: 497). It means that even if a reason for a certain judgement is false, and so the contents of that supported judgement are absurd (in light of which the disbelief or not caring seems to make sense), the paradox still holds. And these features, as well as the aforementioned basis of the paradox – incoherence of attitudes – remind of a certain meta-ethical position. This position is, of course, the RI. And it is exactly the RI that Cholbi thinks can explain the paradox. He himself has given an account of how the paradox arises, what it consists in, of its structure. Meanwhile the RI can give an account of the normal functioning of moral, and normally even wider – practical – judgement making, answering the question of why the paradox arises (and we know the RI account of moral judgement).

I can only remind that it is one of the central internalist intuitions that it would be weird of a person who makes a moral judgement not to be motivated accordingly. Smith, for example, relies on this intuition in his example of giving to famine relief: after my admitting I have a reason to give to famine relief my refusal to do so, without further good explanation, would result in puzzlement on your side (Smith 1994: 6-7). This intuition gives a starting point for arguing in favour of internalism: to insist that it would be weird if internalism was not true. On the other hand, it is a counterpart of the positive intuition that sincere people act in accordance with their words (and reasons).

Internalists share these intuitions with the advocates of Moore's paradox (the ones who believe such paradox to exist in case of practical judgements). Therefore, RI can offer an explanation of the paradoxicality of the moral MP, however, the cases of moral MP cannot offer any independent support to the moral internalism or to the RI more specifically, because the paradoxicality of the moral MP assertions is itself a matter of dispute: not everyone would agree

that (S) is paradoxical at all or at least that it is nearly as paradoxical as (P). But the structural explanation of the moral MP can add to our understanding of the motivational model that internalism suggests, perhaps even change it. So the interesting thing is the possible differences or the sameness of the motivation models that theorists, solving different problems, imply or come up with. And to that we will shift our attention in the next subsection.

Thus, rationalist internalism seems to be a really apt theory for explaining the paradox of moral MP propositions. Cholbi himself would not agree as he claims that rationalist internalists have a wrong understanding of irrationality as pathology. However, as I have already argued, this is a false line of criticism.

One more clarification is due. Cholbi identifies this practical failure as epistemic in character. Epistemic flaw in this sense is certainly different from that which Plato or Socrates had in mind while talking about akrasia. (I want to underline that whatever meanings "irrational" may have, in these lines I will have only the practical irrationality in mind.) Unlike for the ancient philosophers, whether the moral judgement is actually true or not, motivation has to spring, as far as the person remains coherent. Cholbi makes clear: "It is not some objective epistemic fact that establishes whether a given set of considerations is sufficient for an agent to affirm a particular moral judgement. Rather, whatever standards or criteria are adequate for an agent to affirm a moral judgement are themselves adequate for her to be motivated by that judgement" (Cholbi 2009: 503-4). This does not exclude the possibility that sometimes a person might not be motivated in accordance with one's judgement because one doubts one's own standards of decision and sees one's own incoherence with respect to one's other beliefs. But in that case we would say that one does not really make a moral judgement rather than that one fails to get motivated by one's moral judgement.

So the intentions depend on the moral judgements (or normative reasons), so to speak, on what one believes is the case. But as far as the practicing of what one believes is concerned, a moral belief does not have to match the

moral realm without the human being, but only that within. So as far as the practical rationality is concerned, RI is "subjectivist": one has to be motivated by what one believes is true, not by what is – independently of what one believes – true.

Alternative motivational models? It is at least possible to claim that Cholbi's proposed solution for the moral Moore's paradox suggests a slightly different than the standard RI model of the psychological relations in virtue of which the linguistic ones hold. Namely, that the relation between the moral judgements and motivation are mediated by "agent's taking certain facts or evidence as moral reasons for action, reasons that in turn generate motivation" (Cholbi 2009: 506). And this relation, in psychological terms and considering the cognitivist internalist model, is not a direct one between a moral belief and a relevant desire, but that of the believed reasons-moral belief and believed reasons-desire.

To make it even clearer, it seems that for Smith the coherence relation obtains between the moral belief and desire in virtue of a rational person carrying out what she believes is advised by her (hypothetical) more rational self. A rational person knows what desire she must have because of the moral belief which she has, and the belief is normative because of its form – it is expressive of the normative reason that she believes to have. The desire gets its content from the belief and the person is coherent just in case her desire to do something corresponds to her belief. Cholbi's model suggests that the coherence relation obtains between the couples of a reason-belief and a reason-desire in virtue of a rational person's coherence in attitudes to the same evidence. One is coherent just in case one has the same attitude, expressed by desire, to one's believed normative reason as is one's attitude, expressed by belief, to one's believed normative reasons. The desire does not get its content from the moral belief one has; the same content of the belief and the desire is determined by assessing the same reason: both belief and desire are conclusions, supported by the same facts or evidence, expressed by the reason.

Cholbi's scheme shows the same as Smith's: that the relation between the moral judgement and relevant motivation holds in virtue of agent's rationality – because she exercises her capacity to be coherent, to bring coherence into her psyche. However, the difference is the following. The conclusion from the considerations, according to Smith, is established by making a moral judgement. Once it is established, the next step is to desire in accordance with it. It is the moral belief that you should do something in particular that motivates you (also according to Broome). Cholbi, though, lets to see both the making of a moral judgement and getting motivated as the acts of making the right conclusion.

To make a moral judgement, in any case, is to acknowledge there being legitimate reasons for doing something, thus, to make the conclusion of what is to be done. Whether this action, in a rational person (i.e. a person exercising one's ability to be coherent), is necessarily followed by or necessarily simultaneously issues in relevant motivation, is a minor difference. It is not obvious that normative belief, expressive of moral judgement, is a necessary cause of respective desire; it is possible that both of them have a common necessary cause – belief, expressive of normative reason.

The (possibly) alternative explanation seems to at least remotely remind us of the so-called practical syllogism, discussed by Aristotle. The conclusion of the practical syllogism is an action (or act/acting, to be more precise, as strictly speaking action is a full description, and act – just a part of it). In the light of it, the alternative model does not seem to be so odd after all: if making a practical judgement is making a judgement with two upshots – what to believe is to be done (what is the right thing to do, or what ought I do) and what to intend to do (what shall I do), then it means making two conclusions at once. This is, of course, if we think about moral judgement in the cognitivist manner.

Which model is better, is hard to tell. Allen Coates in his paper (2011) argues for one possible advantage of the alternative model. (However, I note at once that I do not find his argument persuasive, so the advantage will not be

established.) Coates's primary goal is to argue for a more subtle formulation of the Enkratic requirement. According to him, it is not enough to say that rationality requires to be enkratic, that is, to intend to do what you believe you should. One can be enkratic and still irrational, say, if the relation in question is due to sheer coincidence. So an improvement to the requirement should secure the non-accidental relation. Coates mentions Smith (together with Davidson) as suggesting to secure that relation by a "because": "it is not enough that you intend to do what you believe you should; rather, you must intend to do it *because* you believe you should" (Coates 2011: 323). However, Coates believes that this answer cannot secure the relation in certain cases.

Coates gives an example, which is supposed to expose a disadvantage of Smith's view. It goes as follows. Ryan believes he should go on a diet and exercise in order to improve his health, but being lazy, he fails to intend that. But when his boss tells him about her new diet and exercise regimen (with a hint he should try it), he, having a tendency to be obsequious, believes he should resist the suggestion. Yet, he begins the diet and the exercising. His belief that he should exercise and diet is formed in response to considerations concerning his health, however, his intention is formed as a response to his considerations concerning his boss. Ryan is enkratic just by chance. Can the "because" of Smith (and Davidson) capture the failure and show that Ryan is not truly enkratic, or that he is akratic?

Coates believes that it cannot. He claims that it would, only if there was a plausible explanation why one should adjust one's intentions to one's normative beliefs. Such an explanation, according to Coates, would be that the beliefs are expressive of reasons. But practical reasons do not consist in a fact that you should do something, but rather in why you should do it; there is a difference between reasons and beliefs. So another response could be such that "responding to practical reasons consists first in forming a normative judgement, which then guides the formation of an intention. On this view, your normative beliefs mediate your intention's response to (what you take to be) reasons" (Coates 2011: 325).

I believe, however, that Coates' attempts to overthrow the latter model of explanation are not successful. Smith can explain why one should adjust one's intentions to one's normative beliefs. The person should intend to do what one judges she should do, because the judgement is based on reasons, it embodies an authoritative directive – you can think of it as of an advice of the one who is better placed to assess the situation, the more rational you. So he handles the supposed counter-examples as well as Coates himself. Ryan is not enkratic as he got motivated not because of his belief that is it right for him to exercise and diet.

But Cholbi/Coates model can also explain it: Ryan was not enkratic because the reason for intention and the reason for normative belief were different. Ryan did not treat his health considerations coherently as he thought they were sufficient for believing that he should exercise and diet, but not sufficient for him to desire to do it; and to the contrary with other considerations concerning his boss. Thus, Coates suggests the same model that Cholbi talked about. His requirement is called that of Coordination: "Your belief that you should do A and your intention to do A are coordinated if and only if they respond (i.e., are based on) the same considerations" (Coates 2011: 329).

So I lay out the possibilities for the psychological model of motivation, but, lacking conclusive evidence, I leave it open which of the alternatives is superior.

5. Autonomy, normativity and rationality

This part of text will be dedicated more to the exegesis of the RI rather than its defence from the criticisms. I have promised that I would finally tackle the question of the superiority of morality, or the claim (1) in its *all things considered* reading: "It is conceptually necessary that, for any action ϕ and any person S, if S judges that it is morally right for her to ϕ, then S judges that she has an all things considered normative reason to ϕ". I have to admit that neither

Smith nor Korsgaard defends firmly that moral reasons should necessarily be overriding with respect to other kinds of reasons. The RI seems to be primarily and essentially a more general internalist position regarding the necessarily motivating character of normative judgements, and moral judgements are a part of them. The categorical character is rather granted to all judgements based on normative reasons, whereas the strength of those reasons and judgements is to be determined by a rational person's preferences. Those preferences are surely restricted by the rationality requirements, but their particular order is not directly and *a priori* dictated by rationality – with the implicitly or explicitly stated probability that morality will come at the top of the preference list. Here we have two main positions.

Smith seems to leave it to be resolved in *ethical* arguments, whether moral considerations should trump all the other. However, at least from Korsgaard's perspective putting morality first is a recommendation that is implied in her arguments. I will come to answering the question superiority of morality gradually, at the same time analysing the remaining essential elements of the RI, such as the source of normativity, the other necessary element (beside coherence) constitutive of rationality, its conception and a possible misconception, the value of rationality and the relation of rationality and morality.

What can close the Open Question (OQ)? As mentioned in Part I, Section 2.5., Rosati traces the force of the OQA to our human nature. And the aspect of that nature that gives it force is the capacity for autonomous evaluation and action, or agency: "The force of the open question argument has deep roots in our agency; it is not merely a function of the expressive and recommending functions of our evaluative notions" (Rosati 2003: 506-7). She claims that if our approval and disapproval were hardwired, then they would not be altered by reflection and the OQ would look to us hollow. Rosati invokes Frank Jackson's remark, where he says that after arriving at mature folk morality to still doubt that the right is that which occupies the rightness

role, would be "nothing more than a hangover from the Platonist conception that the meaning of a term like 'right' is somehow a matter of its picking out, or being mysteriously attached to, the form of the right" (Jackson 1998: 151). I see it as defending the view that the OQ would be futile and so there would be no sense to foster it, if in the end there was no reasoned way to dissolve it, in other words, to close it by a positive answer. If reason has no place in forming a satisfying answer, then it is not reasonable to maintain the question. However, to many theorists (as well as from the common-sense point of view) the question does not look hollow and there seem to be ways (or at least one way) to close it – even if those ways are not reductionism-friendly (as already argued in this text, Part I, Section 2.5.).

Another thinker that is invoked in this context is Hare who also sees an intimate connection between the prescriptivity of moral language and human freedom in thinking and acting: only those who are free need prescriptive language (Rosati 2003: 507, n. 46). Apart from Hare's specific understanding of prescriptivity of the moral language, Rosati embraces the same views. Similar are Korsgaard's views on the matter: people are not determined to act on their strongest inclinations and neither do particular facts suggest certain ways of action. For example, the fact that two towns are 100 kilometres from each other does not mean one should drive those 100 kilometres – this information may be irrelevant for one's goals.

These authors suggest that the OQ arises in virtue of the autonomy of humans, thus, what can close it is an account of good that is rooted in same autonomy. As Korsgaard holds, that which is the source of a problem may also be the source of its solution. In order to close the question, goodness (or rightness, for that matter) cannot be something from an idiosyncratic motivational system of a particular agent. To close the OQ, it should be something that *everyone would be prepared to accept* as a genuine good. I say "prepared to accept" because it is in the nature of being autonomous to be able to accept or reject any proposed course of thought or action. In face of the plurality of values, that definition or analysis of goodness should be very thin.

Perhaps, something that enables the plurality of values or valuing in the first place[94]?

Rosati believes that the goodness judgements engage persons "via the effective operation of those motives and capacities that render them capable of self-governance", that is, they "capture an essential reference to agency and autonomous evaluation" (Rosati 2003: 520). She claims that these motives and capacities differ from others in their non-arbitrary character: without them we would not be evaluators and agents at all. So, one obvious solution to the problem seems to be defining goodness in such a way that it would be constitutive of agency.

Autonomy in thought and action (we will mainly be interested in actions here) is tightly connected to reflectivity of human beings and to normativity of reasons for thought and action. According to this line of thought, the question "what to do?" or "what is good/worthy of pursuit?" arises from the human condition of being a reflective being: asking such a question already indicates there being a distance from one's immediate urges. Such a question supposedly does not arise to beings who are directly guided by their instincts; reflectivity is a necessary condition for autonomy. But if instincts do not perform the function of guides to action, if (the possibility of) being autonomous consists in not being compelled to have a determining power, there is a need of a reliable guide fit for an autonomous person: one needs to freely acknowledge such authority of somebody/something, or to grant it to somebody/something. Thus, a reflective being needs a reliable authoritative guide – a trustworthy, wise advisor. It is no accident, then, that the theorists we talk here about invoke the model of advice.

[94] Korsgaard and Rosati both think that despite the contingencies in our life (what we – to a certain extent – choose as our identities, our values, what choices are open to us), there is a necessary core that enables us to choose values and identities. So if something is to be valued at all, it is the source of the values itself. Rosati: "Unlike our other motives and capacities, our autonomy-making motives and capacities are not arbitrary but, rather, make self-governance possible: they are motives and capacities without the effective operation of which we would not be agents and evaluators at all" (Rosati 2003: 522). I will talk about Korsgaard and this sort of argument for valuing the source of values in more detail further on.

Smith and Rosati suggest us to look at how it works in our lives: when we do not know what to do or when we think ourselves to be bad judges in some situations, we often turn for advice to someone who is wise in general or wiser than ourselves in those particular situations, who is trustworthy and, preferably, knows us. And who can be better placed that our idealised self? It knows our strengths and weaknesses, our dreams and goals, it is the version of ourselves that we strive towards. Humans, not being permanently and reliably guided by their instincts, are not "made," they are the choosers and the makers of themselves in complicated human societies – at least to a certain extent. And so they choose or create their ideal identities that they try to approximate to. "Choosing" or "creating" of the identities only points to the principled nature of their coming into being and the principled possibility of their adoption, as we know that in many cases people get those ideals in more or less ready-made forms from their social environments and circumstance.

In each case, then, a human may consult one's idealised self. Rosati puts it as follows: "something X can be good for a person A only if A would care about X for her actual self, were A under appropriate conditions and contemplating the situation of her actual self as someone about to assume her position" (Rosati 1996: 303-304). Idealised, though, does not mean "perfect" or otherwise out of reach of our actual selves. The idealised self is just a wiser version, that "me" who judges in a cool moment and not in the heat of passions or tormented by afflictions.

For Smith, a normative reason amounts to such an advice that our idealised selves would give to our less-than-ideal selves, or what they would desire our less-than-ideal selves to do in their actual, or less-than-ideal, circumstances (Smith 1995). Normativity, reflectivity, autonomy and the advice model do not connect accidentally – they connect through the structure of two elements. Normativity has a dual structure: it can be born only where there are two interacting parts, where there is a gap. Normativity can be thought of as authoritativeness. Normative is that which is acknowledged as

authoritative; for example, an authoritative answer to what is a good course of action for one's self.

As Korsgaard reveals in her work, normativity has the dual structure in principle. There just have to be two: one that legislates, or governs, and the other one that obeys the laws and can disobey them. There has to be one that could acknowledge or reject the authoritativeness and the other one who could earn the credentials and would have a power to sanction the disobedience. The authorities among people and other animals are common, but humans are special. According to Korsgaard, it is the human capacity for self-consciousness that enables people to govern themselves: we can have the two *inside* of us. The acting self can obey or disobey the edicts of the thinking self. Korsgaard uses the Platonic analogies of the state and a human soul pervasively to highlight this idea and to bring out the necessity of the inner unity for acting.

The idea of authority is different from that of power. Korsgaard explains the distinction in terms of irresistibility: reflection does not have irresistible power over us, but we think we ought to do what we decide to do on reflection (Korsgaard 1996: 104). She says the acting self to concede its right to government to the thinking self which tries to govern as well as it can and which punishes with remorse, regret and repentance. And it is this authority that is the source of obligation (ibid.). Certainly, it is the formal aspect of normativity, its very structure. Thus, what is needed to establish it is a certain relation between the two, based on recognition of desert rather than on irresistible power.

So according to theorists like Korsgaard, it is correct that goodness has to be appealing, that in the end people should want to do that which is good for them. However, many accounts that try to identify goodness with some particular property that people supposedly cannot fail to want, fail. In those cases critics of such theories note that even if goodness must be appealing, appeal is not a sufficient condition: goodness should be appealing in a particular way.

There are some naturalist theories that construe goodness as at least partly constituted by some kind of input from the agent, such as pleasure, want or similar. However, according to Rosati, such hedonistic naturalism (as she calls it) shares a fault with other forms of naturalism that fail to incorporate agency into the account of goodness. That fault is treating as normative natural forces or tendencies that lack normative credentials – even if hedonistic naturalism, contrary to other forms of naturalism, looks for the source of normativity "within an individual's own constitution as a natural and cultural organism" (Rosati 2003: 509). On hedonistic naturalist accounts good depends on a person's current motivational system. And in view of what was already said about the structure of normativity, one can see that in such a case the good cannot be normative: there has to be a gap between the actual and the ideal. For naturalists the good still has nothing to do with person's reflectivity and agency. Rosati notes that in the way the direction of evolution has nothing to do with the agent, her desires may have nothing to do with her as well – in so far as they lack her reflective endorsement (Rosati 2003: 510). Desires, wants may be experienced like a pressure, a violent power from outside of us, unless we embrace them (e.g. all of those cases of an unwilling addict, of a compulsive obsessive etc.). All forms of naturalism, thus, eschew reflective agent from formation of a person's good. The most one can do, according to them, is to learn – by way of discovery or through perception of some kind – from nature (narrowly conceived) what is good for one.

So if normativity is to be conceived not as a natural power, but as an acknowledged authoritativeness, the naturalist theories as mentioned above cannot offer an account of goodness which would incorporate normativity. But that is only to be expected: it is in the very nature of naturalist approaches to understand the sources of morality as rooted in nature where a human is just an instance of it. And it is to the contrary with the constructivist accounts for which morality is rooted in a split human nature. Whereas in the former type of theories reflectivity is understood as an idle human capacity to see the nature unfold through humans, in the latter type of theories reflectivity is a creative

human capacity, expanding the range and complicating the understanding of human nature.

In naturalist theories, the division of authority and power collapses as a person cannot fail to know what is good for one's self, if one knows what one wants, desires or so; unless one is self-deceptive. And the good does not have any critical potential from the person's own point of view: good is that what she wants; there is no space for self-criticism that could issue in changes within the current motivational structure. But that is a lesson that many theorists have learnt and have tried to resolve by introducing, in one way or another, a gap between a person's actual motivational system and the source of the good. For example, Frankfurt's higher order desires are to solve such a problem (the unwilling addict desires drugs, but he desires not to desire them). And each of such theorists needs to justify granting authority to that particular source of good (why the higher order desires are more authentically representative of a person's good; why is it reason or categorical imperative that should be granted the authority; why an idealised self?).

It seems that the OQ is exactly about whether that which claims authority indeed deserves to be acknowledged, or if that which is being universally recommended deserves the recommendation.

The relation of autonomy and orthonomy. But before turning to this, we should tackle the source of one possible misunderstanding: it may seem that Korsgaard and Smith, which in my work represent the RI, are talking about two different things, because while she is referring to autonomy, he uses the term of "orthonomy". However, I want to show that actually, when referring to autonomy as an ideal, or virtue, they both mean the same thing, even if they use different terms, and that is so because such terms like "rational", "autonomous", "human" may have two readings.

These readings could well be described in Aristotelian terms of potentiality/potency and actuality/act: roughly, a human being has a potential to ascend to an ideal of a human being by making actual that which defines

him as a human being. Aristotle does not come up in this context accidentally, for Korsgaard tackles Aristotelian notions in (2008c), (2008b) and elsewhere, and argues for some important similarities of Aristotelian and Kantian accounts of action in (2008e), in (2008d) and in (2008a). And she approvingly incorporates this Aristotelian-Kantian understanding of action into her own philosophy believing that their fundamental notions did not lose their credentials or that some of them can be successfully replaced by equivalents more acceptable to contemporary persons.

Korsgaard notes that terms "reason", "rational" and their Greek equivalents admit of both descriptive and normative use. According to her, the descriptive use refers to activity that can be performed well, badly or not at all (Korsgaard 2008c: 144). However, the normative use describes only an activity done well. So in a *descriptive* sense we *may be* said to be rational, even if we *are not* always such; we *are able* to be autonomous, and so we *are* autonomous regardless of whether we act autonomously at some particular instance; also, we act for a reason even if that reason *is not* a *good* one. But in the *normative* sense we have an ideal in mind, so we qualify something, say, as a reason only if it meets criteria for good reasons, thus, a bad reason in the descriptive sense is not a reason at all in the normative sense.

One could think of description as categorisation in virtue of the characteristic shape or structure, or function[95] of the thing – despite of the defects in it, and the normative usage as a stricter (evaluative?) description that does not include defects, but measures the thing in question against the ideal of

[95] Whether we can always talk about functional definitions of the various objects is not apparent. For example, as far as human made artefacts go, it is quite easy to define them in accordance with their function, but what about animals, plants, people and other living organisms? One way to think of it is in functional terms (e.g. Aristotelian form with a built-in teleology), another is to define them based on what distinguishes them from the others, e.g., what their uniqueness as a species or within the species consists in. Korsgaard in (2008c) supports Aristotelian line, but one does not have to necessarily go that way.
However, I need to emphasise that talking about the normative usage of the various terms it does not imply that normativity of all those notions is of the same type. I.e. the normative usage of "a lung" may have rather different implications than the normative usage of "a reason."

that thing and finds it fit perfectly or not at all. For example, we could say that this is a house (because it is most similar to houses and not to other things), or that it *could* be a *proper* house if it had its roof mended and a wall rebuilt or built, but it is not a *good* house, or not a *proper* house to serve as a shelter for people. Or it is a hound, even if without one of its limbs and a tail, but it is not a good hound that serves its typical functions, or, one could put it: it is not that hounds usually are in this shape, so it is not a hound in a (normative) sense. And so on. Surely, we can use more than those words that Korsgaard has pointed out ("reason", "human" etc.) in the two ways, and so we can see from the examples above (also: "A dog? That rat-like thing over there? That's not a dog! It's ridiculous!"). But we will mostly concentrate on the terms that are the most relevant to us (as in "He's not a human being, that monster!").

I believe it is mainly because of such an ambiguity of word usage (the possibility of descriptive and normative readings) that Michael Smith is criticising the ideal of autonomy (self-rule) and proposing orthonomy (the rule/law of the right) instead. Smith claims that "freedom is not a matter of autonomy, not a matter of being a law to yourself, but rather of *orthonomy*, a matter of having the capacity to be ruled by the right as opposed to the wrong" (Smith 2004a: 109). "Orthonomy" in this case seems to be the normative variant of the descriptively read "autonomy" (Pettit and Smith say orthonomy to be a name for a certain "virtue", 1993: 589). Smith probably allows self-rule to be the rule of the wrong as well as the rule of the right. This interpretation would be supported also by the already discussed Smith's conception of action: he believes that even if instrumental rationality is not all there is to rationality, but it is enough for a person to act instrumentally rationally in order for him to be considered an agent. Autonomy is a part of an instrumentally rational action, but Korsgaard has bigger requirements for action, autonomy for her is associated with the categorical imperative (not just the hypothetical imperative which yields instrumental rationality).

Orthonomy, on the other hand, is a *product* of "our capacity for rational agency" and consists of *two* parts: having the *right* beliefs about what it is

desirable to do and having the *right* desires (Smith 2004a: 109). So in so far as the relations of orthonomy and autonomy are concerned, I think that Smith and Korsgaard only use different terms for the same ideal. As for the agency (along with rationality and normativity) Korsgaard follows Aristotle in her understanding of human agency that springs essentially from exercising one's orthonomy.

An instrumentally rational act for her, as well as for Aristotle, does not suffice to qualify its author as an agent[96]: even incontinent people can be instrumentally rational, but they will not count as rational agents because they are not really *choosing* what to do, they are not doing it for the right reason. In other words, one's instrumental rationality is not enough to qualify that somebody as autonomous. Korsgaard puts it like this: "The incontinent person's action does not count as *chosen* because he does not take it to be worth doing for its own sake; he just wants very badly to do it" (Korsgaard 2008c: 147). Smith puts it similarly: orthonomy consists in forming one's own desires according to the right sort of principles (*orthos logos*), "being sensitive to the properties that count for you as values and not being disrupted by pathologies of desire" (Pettit and Smith 1990: 588). So the problem of an incontinent, or – more generally – of a heteronomous person, is that s/he is not a "more or less consistent executor of such and such a value system", what s/he values does not figure consistently in determination of what to do (Pettit and Smith 1990: 588). For these theorists, there is more to rationality than

[96] "Aristotle certainly does not mean to deny that incontinent people sometimes engage in instrumental deliberation about how to satisfy their unruly passions. … the incontinent person, who does not act from choice, may also deliberate, and in one sense (but not the sense needed for practical wisdom) deliberate correctly" (Korsgaard 2008c: 146). Smith: "orthonomy is the capacity to act in accordance with our normative reasons, a kind of rationality which, as we explain, is distinct from mere instrumental rationality" (Kennett and Smith 2004: 74; also in Pettit and Smith 1990, 1993; Kennett and Smith 1994). Korsgaard: "Now if your will were heteronomous, and pleasure were a law to you, this is all you would need to know, and you would straightaway do act-A in order to produce that pleasant end-E. But since you are autonomous, pleasure is not a law to you: nothing is a law to you except what you make a law for yourself" (Korsgaard 2008g: 109).

instrumental rationality, and we need more rationality than the minimal rationality to recognise the real, or proper, autonomy in action.

That a person is in control, rules one's self, is traceable through one's coherence: if there is a coherence relation between one's value system (what is considered by one to be worthy, desirable, etc.) and one's actions. In other words, if one acts on the right principles – not the principles of *pathos*, but on the principles that one oneself approves of, and that – we can go further – define one's own person. So Smith and Korsgaard agree with regard to this point, as Smith also says that what is wrong with heteronomy, to which he opposes orthonomy, is that desires (which are part of the action, according to him) of a heteronomous person are not expressive of her values (what one would desire were one fully rational). But according to Smith, the important thing for freedom is not only to be ruled from within, endogenously (autonomy as self-rule), because that is not enough, so the problem of heteronomy is not that it is exogenous rule[97]. However, I claim that his account does not clash with neo-Kantian theories, because for Korsgaard, as well as for Smith, what makes for the wrongness of heteronomy is ruling one's self by the wrong/inappropriate laws/principles.

It is possible to see the clash between Korsgaard and Smith, though. For Kantians, the true self is mostly confined to the reason, the active dimension of the mind, and the rest is left to the "outside", the "nature" or to the "alien". It would seem then that what is essential for counting as a non-heteronomous being, is acting on that which is truly your own, not alien, i.e. on the edicts that have their source in reason and not in desires. So it is often thought that desires have no place at all in a Kantian scheme of deliberation, but that is simply not true.

As Korsgaard presents a Kantian picture, often it is inclinations that present a primary proposal for action, and inclinations are grounded in

[97] "We see the non-heteronomous agent, the agent who is practically rational in the narrow sense, as someone in whom desire is appropriately governed, not just as someone in whom the government of desire is exercised by her" (Pettit and Smith 1993: 76).

incentives – features of the objects of those inclinations that make the objects attractive, say, pleasantness. As the will chooses actions (not bare desires), so instrumental reason presents a further proposal for action by adding the means to the wished end, and only then the will determines if it can will a maxim of such an action (such a means to such an end) as a law for itself (Korsgaard 2008g: 109). Besides, for Kantians, *not all* impulses and desires arise immediately from instincts: some are rather a result of reason's work on the material supplied by the instincts, thus, some of them arise from "a complex interplay of instinct and reason" (Korsgaard 1996: 239). The latter explanation makes some desires not as clearly part of "nature" and of "the outside"[98]. However, Korsgaard emphasises that it is not important if desires seem alien to the person or seem to herself to be her own productions, what matters for counting as an agent (and so, for counting as autonomous) is identification with one's principle of choice[99]. I can add: "with the right principle of choice", because the wrong principles are to be excluded by the (correct) procedures of choice, i.e. by the categorical imperative. For Smith the right principle is also a product of correct procedures of deliberation. So I maintain that Korsgaard considers the same: what is important is to act on a right principle.

"Orthonomy" is a term apparently borrowed from the Aristotelian tradition and the latter centres on the so-called executive virtues as well as substantive virtues. Smith makes clear how he understands the distinction of the two: substantive virtues require an agent to be *a lover of the good*, while

[98] Also she claims: "A value, like everything else, is a form in a matter. In the case of value, the form is the form of universal law, and the matter comes from human psychology: some desire, interest, or taste. In that sense, we can see our values as depending on our desires: the objects of desire, ultimately, provide the matter for our values. But it is only the most primitive and basic of our desires that we regard as mere brute likings and dislikings. Values are human creations, but they are not created *ex nihilo* with every action" (Korsgaard 2009: 209).

[99] Because of such an acknowledgement Korsgaard allows even for identification with principles of treating one's desires as reasons: "If the law is the law of acting on the desire of the moment, then the agent will treat each desire as a reason, and her conduct will be that of a wanton. If the law ranges over the agent's whole life, then the agent will be some sort of egoist" (Korsgaard 1996: 99). These principles of choice, though, need to meet certain requirements, imposed by the categorical imperative.

executive values, such as temperance, courage, fortitude and justice, require one to be *a good lover* (Pettit and Smith 1993: 76). It is also said that the latter virtues are requirements of orthonomy.

Korsgaard in her study of Aristotle notes that acting on *orthos logos* includes doing "the right act at the right time in the right way *and for the right aim*" (Korsgaard 2008c: 146). But Korsgaard also is arguing for a similarity of Aristotelian and Kantian understanding of a good action, and that what for Aristotle is – let us use the term – an orthonomous action, that for Kant is an autonomous action. For Aristotle, acting on *orthos logos* means choosing to do a certain act-for-the-sake-of-a-certain-end for its own sake (and because it is noble) (ibid.: 147). Such a choice is not just instrumental, it is more than that because it is the whole combination of the means for a certain end that is chosen given *the whole combination* is seen as worthy of pursuit in those particular circumstances. From Korsgaard's Kantian perspective, what enables a more than instrumental choice is categorical imperative of which hypothetical imperative is only an inseparable part (see Korsgaard 2008h). A good, or moral, action is an autonomous action, or an action in sync with the categorical imperative, i.e. chosen because it is intrinsically good, has the form of the law. "Aristotle thinks a good action is one whose agent sees it as the embodiment of right reason, just as Kant thinks that a morally worthy action is one whose agent sees it as an embodiment of the very form of law" (Korsgaard 2008d: 191). So both a noble action and a good-willed action embody a principle of reason (ibid.: 193).

Rationality as coherence and autonomy/orthonomy? If we take coherence to be a necessary part of the conception of rationality, we can ask if coherence is also sufficient for it. Say, if a person accidentally acquires the desire to do what one thinks is right to do on a particular occasion or accidentally makes the correct conclusion with regard to the evidence presented to her, would we hold such a person rational? It seems that being rational is not just something that happens to one, rather, in order to be called

rational one needs to have been active with regard to one's decision or action at the moment when it takes place, and not, say, if one retroactively conjures up a rationalisation. One can even say that it is already implied by the very concepts of "decision" and "action". The coherence at hand, thus, is the coherence actively brought about, not a happy accidental occurrence of coherence within a person. Remember an externalist position that talks about the need of distinctly moral desires for a moral motivation to take place. On their account a coherence relation between a moral judgement and a moral desire obtains as well, however, this relation is not actively brought about by reason.

Korsgaard, for example, contrasts two cases. In one, a person is conditioned so that when wishing to drink and believing that the object in front of him is a pencil sharpener, he wants to put a coin into it. In the other, a person rather puts a coin in a soda machine. The one action is mad, the other is rational, but the difference, according to Korsgaard, is not that in one the relation between the belief and desire is of the wrong kind (pencil sharpeners are not sources of drink) in the other – of the right kind. The latter explanation concerns only the relations of the belief and desire, making a rational action accidentally different from the irrational one: "After all, a person may be conditioned to do the correct thing as well as the incorrect thing" (Korsgaard 2008h: 33). However, Korsgaard adds, "the correctness of what she is conditioned to do does not make *her* any more rational" (ibid.). So rationality must include the aspect of activity; the motivation has to spring from a person's own recognition of the propriety of the aforementioned relation or she has to combine the two elements appropriately: "A person acts rationally, then, only when her action is the expression of her own mental activity, and not merely the result of the operation of beliefs and desires *in* her" (ibid.: 33)[100].

[100] Korsgaard herself notes the similarity of this requirement to another Kantian requirement concerning duty: "One way to put the point of this paragraph is to say that a rational agent must act not merely in accordance with reason but *from* it. … The debate between the rationalists and the empiricists about rationality could then be constructed as proceeding in the way their debate about the relative merits of acting in accordance with duty and acting from it actually did" (Korsgaard 2008h: 34).

Whether it is an action ("theoretical", that of arriving at a certain conclusion/decision, or "practical", that of arriving both to a certain decision and of getting motivated respectively) or a person that we attribute the predicate "rational" to, it is a deserved predicate, attribution of a non-contingent coherence relation. This means that to attribute rationality to someone is to attribute responsibility[101] for that thing (decision, action) in virtue of which one deserved this predicate. Thus, we can discern the elements of activity, necessity, connections to responsibility.

It is no wonder that we have to add this. It is evident that what we attribute rationality to is either people (I will not take a stance on the possibility of non-human animal rationality) or actions/beliefs/decisions/choices. People can be rational insofar as they are capable of being active, thus, of bringing about changes in themselves or in the outer world. We, after all, do not attribute (ir)rationality to natural processes or functioning of human made devices, nor to the natural processes within a human. Thus the link between rationality and human activity, or agency. It is not surprising also because rationality is an attribute of the workings of reason – the faculty that is "the active rather than the passive or receptive aspect of the mind. Reason in this sense is opposed to perception, sensation, and perhaps emotion, which are forms of, or at least involve, passivity or receptivity" (Korsgaard 2008e: 2).

This additional necessary element of rationality is evasive of naming. Smith approaches it with the term of "disposition towards coherence": "[rational people] do not just so happen to care about being coherent – something an equally rational creature may just so happen not to care about – rather, being rational is, inter alia, a matter of being disposed to restore

[101] I am not going to investigate questions of responsibility in detail, because it is a complicated matter. I just note that people may also be responsible for irrational actions, however, these have to be actions in their characteristic shape, even if they are not perfect actions, if they are failed actions or bad actions. In other words, to say that someone did something irrational does not revoke that agent's responsibility: it is as if saying that one has failed in acting rationally. The relevant distinction is rather that of somebody acting (bad or good choice) *vs.* something happening to somebody (no choice), where in the latter case one is usually not responsible, unless there are complications.

coherence: the disposition towards coherence is partially constitutive of being rational" (Smith 2004a: 97). That capacity, at the end of his (2004a) paper, turns out to amount to a capacity to be free, or orthonomous. For Smith (and Pettit), orthonomy is a version of the ideal of autonomy.

Korsgaard defends autonomy as a necessary part of rationality as well. According to her, in the Kantian tradition, of which she is a part, "to be rational *just* is to be autonomous. That is: to be governed by reason, and to govern yourself, are one and the same thing" (Korsgaard 2008h: 31). I should note that as rationality is a predicate applicable not only to actions, but also to beliefs, so freedom, or autonomy, is not only possible as a freedom of will, but also as a freedom of belief (even if most of our attention will be given to the autonomy of the former kind).

Self-government or self-control, self-determination, freedom of will and the related question of responsibility are big and contentious topics that are dealt with mainly in other branches of philosophy as ontological questions and their practical implications. And I am not going into these discussions. I will only note a couple things in this respect. Adherents to the RI believe that freedom is a necessary supposition not just for moral thought, but for agency (basically) and our practices (not just moral ones) as such. One can defend this supposition on different grounds, though.

On the one hand, the approach to ethics that I analyse and sympathise with treats ethics as a practical enterprise, not as a branch of epistemology. As Kant believed, the supposition of the possibility of free will is necessary for moral philosophy and moral life, thus it is supposed. Changing the supposition would mean changing topic, or passing to the "theoretical" perspective where people are viewed third personally as determined, and so where "choice" is not a part of the vocabulary. On the other hand, one can assume an even more cautious position such like Smith's (and Pettit's), claiming his enterprise to be purely conceptual: the spelling out of our concepts of, in this case, freedom of will and responsibility. It is a different question if these concepts denote any reality and if so, to what extent. However, Smith and Pettit stress that it is

possible to reconcile this supposition with the scientific knowledge of the world.

For people, being condemned for the first personal stance, means being condemned for the idea of freedom – whether illusory or not. It seems inevitable. However, if one can conceive of everyday practices not "biased" by the supposition (as if the "lenses" that we thought to be a part of our eye were discovered to be possible to take out), then we could or even should give up some of our practices along with the supposition. As Smith and Pettit demonstrate, our interpersonal and even intrapersonal (meaningful) conversations are grounded on the same supposition of ourselves and others (at least in most ordinary cases) being free and responsible believers or desirers. The idea is this: if we do not believe that people are able to recognise the norms governing thought and action and to respond to them, that is, if we think of them as incapable of effective deliberation, then there is no point for us in conversing with them on matters of thought and action, taking others and one's self seriously (Pettit and Smith 1996: 440). The suppositions being shed, the practices of conversation would become futile as a means to attain certain goals, such as to get to know what the case is or what the desirable course of action is, and taking others' views seriously.

The change of these practices would mean that people "would have to adopt a wild and self-defeating stance on one another and on themselves. They would have to discount everything they must assume in order to practice conversation, and relate more broadly in an interpersonal fashion. Indeed, since thinking itself is a kind of intrapersonal conversation, as we saw earlier, they would have to discount everything they must assume in order to practice conversation with themselves: everything they must assume in order to think" (Pettit and Smith 1996: 447). Thus, the two authors believe people rightly treat each other and themselves as free and responsible and that this "conclusion is inscribed in habits of thought that we can scarcely imagine anyone being prepared to give up" (Pettit and Smith 1996: 448).

One could think that this inevitability of the supposition pushes us to the side of realism. After all, recalcitrance of the phenomena to manipulation by humans usually, even if not reliably, points to the reality of those phenomena, to their not being pure phantasms of our mind. Colour experience is an instance of a similarly real thing. We (people with normal eyesight) cannot stop seeing things as coloured. Bet let us imagine that, knowing why we see colours, we can shift to seeing things not in terms of colours, but in terms of wave lengths. Apart from the aesthetic value of the colour experience, it seems that nothing in their function of providing information that is important for our guidance in the world would be lost – the same function would be executed by the wave lengths. For example, I would know which tomato is ripe and which is not, which traffic light is on, etc. The change would not change my life, the substitute of colour experience would provide me with the same guidance in the world.

Now, let us think of a similar thought experiment in relation to our moral perception. Knowing that feeling free is like seeing colours and knowing how we should instead think of it in "scientific" way, let us suppose we could shift instead to seeing things, including both other people and ourselves, as determined. It seems that in such a case my practices would change, the guidance in the world would be changed along with the "lenses" through which we would look at it.

Acknowledgement, within the sphere of moral philosophy, that freedom of will or thought is an illusion, that people are indeed determined, would not just have the same effect as pronouncing some fact, but would set a certain norm, would give the normative glaze to the descriptive concept that it should not get. It seems to open the same door that Pascal was warning about: "Inequality must necessarily exist between men. That is true, but having granted it, the door is open not only to the most overt domination but also to the most overt tyranny" (Pascal 1999: 114). Freedom is not just a descriptive concept, it carries normative implications with it. As Smith said, "[w]hether

our concepts of freedom, responsibility, and value stand or fall, they do so together" (Smith 1997: 100).

It does not mean, though, that the fear of having to reform our moral, legal and political systems and change some of our everyday practices should sustain a false supposition. However, Smith thinks that the question of whether we are indeed free and responsible (and, if so, to what extent) can only be decided by engaging in a substantive normative debate (Smith 1997: 110). The same test is awaiting the other interrelated concepts, such as "rightness", "moral facts" and so on.

But we can return to the idea already put forward earlier on, namely, that people can be (ir)rational only insofar they are active, when they act, are agents. Here I once again stress that agency is to be understood more widely – a person can be an agent in the sphere of thought and in the sphere of action proper. Making a proper conclusion in an argument is also an action. Thus, rationality presupposes agency. On the other hand, responsibility also presupposes freedom, or agency (or autonomy). The concepts are interrelated. And indeed, "freedom" in contemporary practical philosophy contexts is replaced by "agency". As Korsgaard notes,

> "Many of the problems that are now discussed under the rubric of 'the philosophy of action' were once discussed under the rubric of 'freedom of the will,' and this is no accident. Agency is almost as mysterious as freedom of the will, and for the same reasons – with this important difference: that it is much harder for skeptics, even those with 'scientific' pretensions, to deny that agency exists" (Korsgaard 2008e: 10-11).

And rationality in the accounts of the adherents of the RI stands at the very sources of agency.

Thus, the evasive element of rationality, beside coherence, is autonomy. Autonomy in a way seems to give more substance to the relation of coherence, to guarantee that essentially necessary relation between a reason and a judgement and between the reason and motivation (or judgement and

motivation). By some, introduction of autonomy into the conception of rationality may be seen as a drawback of the RI. By the lights of some others, it is just part of the analysis of rationality.

6. Morality as the supreme form of rationality?

The value of coherence and rationality. But what is the intrinsic value of rationality? Why should one care about being rational, that is, about being coherent and autonomous/orthonomous? For example, Zangwill calls the value of rationality into question. He admits, as an externalist, that instrumental rationality may sometimes (but not always) be valuable as providing the means for desire satisfaction. However, Zangwill claims that instrumental rationality may also be counterproductive, may take too much time and effort or even fail to help to attain the end (as in case of trying to fall asleep one would better not think of the end, but count imaginary sheep) (Zangwill 2012: 355). He also mentions the cases when we have to rely on our natural tendencies to track an end or means, as when catching a ball or avoiding danger we should listen to our instincts. Other cases are those of associative thinking, daydreaming, personal interactions, telling jokes, dancing, musical improvisation, falling in love and similar. Zangwill concludes: "Indeed, for most of the best things in life, the faculty of reason does, and ought to, take a back seat!" (Zangwill 2012: 356).

So, first, he argues that rationality is not always an effective means of forming an intention for action: sometimes relying on non-rational forms of thought is preferable. Second, exercising rationality is not the only means of creating valuable experiences in life, or it is not the only source of value. Zangwill cannot agree that rational deliberation should be the only good way of thinking what to do: "But there are good ways of thinking – creatively, associatively, empathetically, imitatively, intuitively, imaginatively, for example – which are perfectly good in their own way" (Zangwill 2012: 356-357). He supposes that the Kantian picture is such that the ideal person is

always in control, and other forms of thought at best must be subject to reason. For Zangwill that is not necessarily an ideal and inspiring picture of a human being, so why should we prefer being rational as supremely valuable?

Korsgaard suggests that the role of rationality is that of the unification of the self: the intrapersonal and trans-temporal coherence guarantees continuity of personality and effectiveness of deliberation (in contrast to an innerly conflicting and unpredictably behaving individual). However, even this does not convince Zangwill as he takes a Nietzschean line of argument that fragmented personalities may be more energetic and creative unlike the dull and uncreative rational and unified souls. So again, why rationality as a superior value? The more so, Zangwill sees a further problem in Korsgaard's approach which grants rationality an existential value: "How can being non-rational sometimes be a better way to be? It cannot, if being subject to the demands of rationality is the essence of a person. If rationality is the supreme value, then the kinds of mental life that Nietzsche celebrates are unacceptably downgraded" (Zangwill 2012: 358). So for him, the biggest problem is that rationality is the supreme value and the supreme form of thought.

One must agree that many good things in life are not due to reason or reasoning and that it is arguable if a rational person is the only ideal to aspire to. It is also true that other forms of mental life are downgraded. However, I want to highlight the bases for this preference of reason in the theories of the Kantian type. These theories claim that we are (at least in so far as we know) distinctively human in virtue of our possibilities and structure of reason which supposedly enables control over ourselves: "the space of reflective distance presents us with both the possibility and the necessity of exerting a kind of control over our beliefs and actions that the other animals probably do not have" (Korsgaard 2008e: 4). And it is the same structure of reason that puts us in need of good tools of self-guidance: if we have the control, we need "landmarks", we need a source of authoritative practical decisions. So it is the same structure of reason that enables us to value something (we are asking what is good and what is right, the world only provides us little guidance in

this search – in so far as we have bodies and certain physical and psychological needs), it is the source of values. That is why this ability of particular part of our mind, the reason, is definitive of us *as humans* and why it is valuable to us, *human beings* (without it, we would presumably not be distinctively human and would not be able to value at all). As Korsgaard says: "You must value your own humanity if you are to value anything at all" (Korsgaard 1996: 123).

It so happens that Kantian theorists also argue that morality is and has to be based on reason, because values in general and moral values in particular, as already mentioned, are not part of nature, but are to be created (even if not in a relative manner) by the same humans out of need. And these theorists believe that giving this gift away, returning it back to nature by choosing to listen to something else than reason would be unfair. But they do not say that only reason is valuable or that human life must all be strictly controlled by reason.

It is true that the theorists we analyse here talk not just about normativity of moral judgements, but also that of other kinds of practical judgements as dependent on reason; also they consider behaviour to amount to action only if it is rational. After all, for such a neo-Kantian as Korsgaard "The categorical imperative … is not just the principle of morality. It is also the constitutive principle of action" (Korsgaard 2008e: 12). In that case, it seems, they require reason to regulate quite a large part of our lives. So it is correct that rationality and reason are seen as a big value – by necessity from the logic of these theories. It is even the supreme value and ability, especially as it comes to the sphere of the specifically human. But not only that – Zangwill is correct saying that other forms of thought at best must be subject to reason. Reason seems at least to provide the "humanising glaze" to the products of other abilities of the human animals. Because rationality just is constitutive of human activity.

Surely, sceptics about reason have a right to retain their position as to the functions of reason and as to what constitutes humanity. This is one of the main differences in position that can hardly be changed by an argument. I just want to emphasise that the importance of reasoning and of reason as such is not an absolute value out of nowhere, but the supreme value in so far as (the

specificity of) humans are concerned because reason may be considered to be constitutive of humanity. But of course Zangwill is right that happiness is not exhausted by being moral or rational.

In addition, we can provide other advantages of reason and reasoning, which may be more independent of the constructivist story about the constitution of humanity, but which would present an attempt to answer Zangwill's question "but why is it important and more important than other values or forms of thought?" I will discuss the value of rationality shortly.

But another important reason for preferring rationality, especially in morality, is the following. It is true that different kinds of thoughts (associative, intuitive etc.) are efficient, but they do not operate in different people in the same ways and do not necessarily yield the same results. The results of these operations of thought, even given the same data, are often subjective and accidental. But the power of reason which is in principle available to every human being (and which actually is often used by most or at least many), which is susceptible to orderly procedures, and which can necessarily guarantee the same intersubjective result if those procedures are followed appropriately, is exactly the exercise of rationality and not of those other forms of thought.

What is the value of rationality as of coherence and autonomy/orthonomy, if we look at the question more independently (i.e. not deriving it from a constructivist logic)? Let us take an example. Let us say that a certain person *A* acts or believes in accordance with one's best reasons for acting or believing something. And another person *B* does or believes something else than what s/he oneself judges that s/he has the best reason to do or to believe. From the third personal point of view, person *A* seems to be trustworthy, having integrity (in both senses of the term – honest and unified), transparent, whereas person *B* is epistemically suspicious. From the first personal point of view, person *A*, if one's actions and beliefs were coherent not accidentally, feels to be congruous and efficient, whereas person *B* is either a

liar or a failure – in which case s/he is innerly divided, in tension with one's self.

Coherence as such does not seem to have much of a value on its own, even given it is not an accidental relation. Let us say that a person *C* judges (falsely) that one has a conclusive reason to exterminate a group of people based only on their skin colour. *C* carries out one's own judgement. Person *D* has the same (false belief), but is weak-willed, so *D* fails to act on one's (falsely) strongest reason. In the face of the latter two examples, the efficiency, transparency and integrity do not seem to be big virtues of the murderous *C*.

The same is on the theoretical level: if someone is coherent in reasoning, we can trust the results of such reasoning, but not in the sense of their truthfulness, because if the suppositions of such reasoning are silly or false, coherence is not worth much. On the other hand, if somebody comes to the right conclusions, but does so accidentally rather than by coherent reasoning, we will probably not acknowledge such a person and such a procedure to be responsible for the results. And so on the practical level: if a person is coherent carrying out stupid or cruel beliefs, we will not consider coherence a particular virtue, but if such a person will accidentally do something good, we will not attribute him or her full responsibility for the good actions.

So we can draw some conclusions. If we observe incoherence in somebody's talk or behaviour, we get a signal that something is not working as it should, we doubt the authenticity of the talk or behaviour. Such a person or result is untrustworthy. The minimum value of coherence is its aspect of trustworthiness (or transparency, or authenticity) and its indication of the possibility of agency (or of effectively functioning unit). The possibility of agency (in belief or action) becomes actuality of agency if we add to the idea of rationality, constituted mainly by the idea of coherence, autonomy/orthonomy.

Rationality and morality. And, subsequently, what is the relation of morality and rationality? Another of Zangwill's problems with the RI lies in

the Kantian reduction of moral norms to rational norms. According to him, these norms are distinct and the normative question for moral norms cannot be answered by pointing out to the normativity of rationality (the more so, of non-instrumental rationality which, by Zangwill's lights, is an implausible kind of rationality). The question "Why be rational?" is to be answered by our desires (or our want to satisfy them); a desire is the source of this kind of normativity. However, moral normativity, according to Zangwill, is due to the existence of moral properties. As already mentioned, Zangwill presents a metaphor of treading through a moral mine-field to make morality seem more relevant than the Mount Everest. Also, he answers the "Why to be moral?" with "Not: 'Because they [the moral properties] are there', in some distant sense, but 'Because they are here' in our very acts" (Zangwill 2012: 362), adding that being moral is not optional.

However, as to the latter point, I have already shown that the metaphor is not apt, and if it is not, then it is very hard to understand why moral properties are so different from other properties in the world. I.e. it is not obvious why moral properties are normative with relation to our decisions whereas other properties are not so or not in the same way. Why does it not spring from colour properties or the property of resilience etc.? Or is all the reality equally normative, authoritative in the same sense? And what does this normativity amount to? Moral properties are normative, but it is not the apprehension or endorsement of that normativity that has an impact on our decisions, but an accident; it is not normativity of rationality that moves us to do something rational, but again – an accident, a desire that cannot in principle spring from a controlled activity of reason. I find this picture weird.

The only thing that works in Zangwill's "Because they are here" is not the idea that moral properties are somehow closer to us than non-moral natural properties, but the idea that normativity is connected to the situations that we find ourselves in in some way or another and that we have to choose inevitably: "You may or may not choose to go to the mountain. With morality, the mountain comes to you" (Zangwill 2012: 362-363). The normativity does

not consist in the proximity or the inevitable character, it is not something that things inevitably have, but it has to inevitably be granted to some things in virtue of how the humans are.

But if rationality amounts to coherence and autonomy, does morality equal rationality, i.e. is morality – a matter of being coherent and autonomous? How does it differ from other kinds of instances of being rational?

The rationalist theories, first of all, care about the form: they do not give substantive definitions of rationality or of morality. So moral judgements or moral reasons may be different from other kinds of practical judgements or reasons in their contents, but not in their form: it is in virtue of that form that they count as practical judgements or reasons. The content of moral judgements and reasons is circumscribed by those formal requirements as well as by the understanding that the other-regarding character is characteristic of morality. Smith seems to leave it open for different interpretations what moral content should be identified by:

> "First of all, we find out what all of our normative reasons are. Next we look to see whether any of these have the peculiar features of moral reasons: that is, we look to see whether there are any normative reasons that are other-regarding, or which require us to ascend to an egalitarian plateau, or which require us to promote human flourishing, or whatever else might be thought to be the distinctive feature of moral, as opposed to nonmoral, reasons" (Smith 1997: 107).

Korsgaardian view is similar. However, I want to emphasise that neither for Smith nor for Korsgaard is moral judgement universal to the extent that the particular circumstances of the agent would not be included into the universalization process. That is, it should not be presumed that a proper moral judgement is so universal that it covers all possible circumstances and that those circumstances are not reflected or are not implicit in the judgement. And the same is to be told of any practical judgement. Thus, these accounts are immune to the usual criticisms advanced to overly-universalist moral theories

that they do not take specific circumstances into account. Also, they are not susceptible to the charge of making every rational person to be the same as to their desiderative profiles. Smith says: "There is no suggestion that fully rational people will all have the same tastes in food, and clothes, and basketball teams. ... The claim is rather that they will all converge in their desires about what is to be done in highly specific circumstances" (Smith 1997: 89). Therefore, such a thing as preference for beers or wine or being the father of a particular child are all relevant circumstances to moral or non-moral normative reasons and judgements[102].

As for Korsgaard, her conception of morality, on the one hand, seems to be very wide, ubiquitous – like that "humanising glaze" of rationality over human decisions, on the other hand, after all, there seems to be something strictly moral.

It is rather natural that rationality and morality have the same source and common features: "Since moral or ethical value pertains only to human action, it seems natural to think that it is somehow related to, or supervenes on, the specific character of human action" (Korsgaard 2008d: 174) as for a human action one has to be rational. The way the action is chosen, a form of an action, seems to be definitive of morality – remember Kant's distinction between an act from inclination (and in accordance with a duty) and an act from duty. So for Korsgaard the fact that some act is done on a principle, on a valid maxim is constitutive of its moral character rather than its purpose. Also, as she says, a distinguishing feature of a moral agent is truly active, authentic actions, not just reactions (Korsgaard 2008d: 175). This is a very wide understanding of morality indeed: it seems that almost every action – despite its purposes – becomes moral in virtue of meeting certain requirements.

[102] Another instance of the balancing of the public and private that this theory offers is that the reasons are understood as something public, having the universalist character, but it is your special relation to that which is publicly good that constitutes the element of privacy. As Korsgaard puts it, "In other words, to have a personal project or ambition is not to desire a special object that you think is good for you privately, but rather to want to stand in a special relationship to something you think is good publicly" (Korsgaard 2009: 211).

Those requirements are expressed, first of all, by the categorical and hypothetical (instrumental) imperatives. On this view, the categorical imperative is not only a principle of morality, but also the constitutive principle of action. Second, hypothetical imperative is inseparable[103] from the categorical one, "rather, it picks out an *aspect* of the categorical imperative" (Korsgaard 2008h: 68). The rationale for such a conception is that in order for someone to be an agent, one needs to be, "at once, autonomous and efficacious – it is to have effects on the world that are determined by yourself. By following the categorical imperative we render ourselves autonomous and by following the principle of instrumental reason, we render ourselves efficacious" (Korsgaard 2008e: 13). These are the constitutive principles of agency. The inseparability of the two principles comes with the thought that categorical imperative is a *practical* principle, and practical principles tell us to *do* something rather than what ends to have or what means to some ends to take (Korsgaard 2008h: 68 and 2008a: 222, n.17).

Korsgaard can unify the two principles also because of the conception of action that she supports and that I have clarified already. So the maxim that is tested by the practical imperative includes both the act and the end of that act: "because the question raised by the categorical imperative test is whether there could be a universal policy of pursuing *this sort of* end by *these sorts of* means" (Korsgaard 2008a: 218). Second, she holds a parallel with Aristotle's theory of the unity of the virtues in relation to the distinguishability of different kinds of practical (ir)rationality: "there is only one principle of practical reason, the categorical imperative viewed as the law of autonomy, but there are different ways to fall away from autonomy, and the different principles of

[103] From 1997 (when her "The Normativity of Instrumental Reason" was first published) to 2008 (when the same paper was reprinted in a collection of her papers with an Afterword) Korsgaard has changed her position to a more radical one: "I now think that what I say about this in "The Normativity of Instrumental Reason," Essay 1 in this volume, on pp. 57–8, where I portray an agent as enacting ends into law prior to enacting means into law, is misleading. At the time I wrote that essay, I believed that its argument showed that hypothetical imperatives depended on categorical ones; as I say in the Afterword to that essay, I now believe it shows that, strictly speaking, there are no separate hypothetical imperatives" (Korsgaard 2008a: 221, n. 15).

practical reason really instruct us not to fall away from our autonomy in these different ways" (Korsgaard 2008h: 63, n. 60).

So categorical imperative as the practical principle constitutes human agency by enabling efficient and autonomous functioning of a unified individual. However, it seems that not every principled action is rational or moral. For example, Korsgaard mentions that some actions may be governed by "a principle of choice which is not reason's own: a principle of honor …, prudence …, wantonness …, or obsession" (Korsgaard 2008g: 117). This principle unifies the person as well, however, this unity is contingent and unstable. Animals also act on principles which just are instincts, they are intelligent, but not rational. So one clear restriction on the principles of rational and moral actions is that they have to be *reason's* own principles (such as categorical imperative).

This restriction is a natural part of the constructivist position of this kind. As mentioned, constructivism thinks a solution to the fundamental human problem to be found in the same source as the problem. According to Korsgaard, as Rawls moves from a concept of justice to the conception of justice, so does Kant – but with relation to the problem of human agency ("what to do?"): "Negative freedom is the name of a problem: what shall I do, when nothing determines my actions? Positive freedom proposes a solution: act on a maxim you can will as a universal law" (Korsgaard: 2008f: 322). In other words, if nothing determines you, determine yourself in a law-like manner. The source of the problem is reason, so reason is its solution as well. One can perhaps doubt if categorical imperative is a "reason's own principle", but I will not get into these debates here.

Another requirement to the moral actions is present as well. Korsgaard makes a difference between the categorical imperative (and instrumental principle is a part of it) and the moral law which is one of the formulations of the categorical imperative in Kant's works, but which Korsgaard sees as having a distinctly moral character. The categorical imperative unites a person synchronically and diachronically enabling her/him to be efficient and to have

an identity. Such a person, as mentioned, may even be a wanton or an egoist. But the moral law, which tells us "to act only on maxims that all rational beings could agree to act on together in a workable cooperative system" (Korsgaard 1996: 99), adds an interpersonal dimension to the requirements of coherence. So I believe that one could understand those actions the maxims of which pass the categorical imperative test as *good in some way or respect*, good for some person or group of people, as rational actions (to some degree), whereas the actions based on the moral law are *good* for every person and all people, those are perfectly rational actions.

Another restriction becomes clear from comparing human action with behaviour of non-human animals. Korsgaard claims that in becoming motivated the idea that your motives are good are crucial[104]. It is the reflexive structure of motivation, its self-conscious character that distinguishes human actions from animals' behaviour: "It is this property – consciousness of its own appropriateness – that the lioness's motivation lacks" (Korsgaard 2008a: 214).

Therefore, one can see that indeed the range of the moral actions is rather wide, even if it is restricted not only by the categorical imperative, but also by other additional requirements, especially the moral law. At the same time, of course, this morality bleeds off: in virtue of not having substantive contents and not being equivalent to substantive values (e.g. such as pleasure), it is more like a constitutive definition of an authentic, ethical way of life. A principled action based on reasons of prudency is held to be by Korsgaard defective, so a moral action, then, is a perfect action, a human action *par excellence*.

Another way to look at the matter is to see how the distinct character of morality is brought out through Korsgaard's conception of practical identity. A practical identity is such a "description under which you value yourself, a

[104] However, one should not think that it is a view that confuses proper reasons for good action with fetishism about self-righteousness or with moral fetishism: I have already presented the conception of Kantian action before – distinguishing the purpose of an act and that of an action. For Korsgaard this reflexivity rather serves to illuminate the difference between Kant's *acting in accordance with* a principle/duty and acting *from* a principle/duty: the one who acts from a duty does not have a different purpose, but is conscious of the good-making properties of an action.

description under which you find your life to be worth living and your actions to be worth undertaking" (Korsgaard 1996: 101). Every person has several practical identities, such as "a human being, a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone's lover or friend, and so on" (ibid.). It is these identities that give rise to reasons and obligations – that is, they give content to the various practical judgements: "Practical conceptions of our identity determine which of our impulses we will count as reasons. And to the extent that we cannot act against them without losing our sense that our lives are worth living and our actions are worth undertaking, they can obligate us" (Korsgaard 1996: 129). One of such identities is the fundamental identity of a human being.

Korsgaard says that not all identities are equally important and equally necessary. Many of them are gained in life accidentally, some are chosen. All of those accidental identities can in principle be shed. However, those conceptions of ourselves which are the most important can give rise to unconditional obligations because violating them threatens our identity (that we cherish) and our integrity. Sometimes it is even better to die than to lose some of them. But according to Korsgaard there is one identity which is necessary, which cannot be disposed of. And it is a moral identity – a "human identity conceived as a form of normative practical identity" (Korsgaard 1996: 125).

One note is due. Even if Korsgaard mostly uses terms of moral and human identities interchangeably, one can see a gap between a human identity (identity of a reflective being) and a moral identity. However, I believe that the possibility of difference of the two identities should be seen as merely a difference of degree or a difference between a potency and act: a human animal necessarily has the potency of being rational because of one's structural characteristic of the mind – reflectivity, and one is maximally rational when one is moral. The more so as Korsgaard sometimes defines moral identity in terms of a certain attitude of endorsement of one's own human identity: "To treat your human identity as normative, as a source of reasons and obligations,

is to have what I have been calling 'moral identity'" (Korsgaard 1996: 129). In other words, what does it mean to be a human? At the minimum, it means to be reflective, to be rational to some extent, at the maximum – to be beyond rational criticism, to be moral[105].

Korsgaard argues that people have moral identity in virtue of being human beings, it cannot be rejected – "unless we are prepared to reject practical normativity, or the existence of practical reasons, altogether" (Korsgaard 1996: 125). It has to do with the reflective distance in the human mind which enables us to choose our beliefs and actions and which makes us choose the proper grounds for those decisions. Being the source of this fundamental freedom, reason should be the fundamental value as well. First, it must be a value: "if you value anything at all, or, if you acknowledge the existence of any practical reasons, then you must value your humanity as an end in itself" (Korsgaard 1996: 125). Second, it must be the fundamental value on pain of incoherence: "In so far as the importance of having a practical identity comes from the value of humanity, it does not make sense to identify oneself in ways that are inconsistent with the value of humanity" (Korsgaard 1996: 126).

On the one hand, we see that according to this account all other practical identities logically depend on the human identity: it makes it necessary to have those others, besides, part of their importance and normativity comes from it. On the other hand, other practical identities have some independence as well. And even if Korsgaard tries to argue for the supremacy of morality, she remains honest. She is well aware that even those who acknowledge their obligations to humanity, are never just moral agents, but see their obligations to particular others as independently forceful (Korsgaard 1996: 128). Korsgaard does not deny the possibility of the conflict between different identities and does not want to "to remove its sting. Conflicting obligations can

[105] Again, in so far as rationality is a matter of degree and is constituted by coherence and autonomy, we can be coherent in different ways. For example, we can have intrapersonal and trans-temporal coherence, but we can be even more rational – up to being moral – by reaching a high degree of interpersonal coherence.

both be unconditional; that's just one of the ways in which human life is hard" (ibid.: 126). In other words, a human being is a citizen of the Kingdom of Ends, but also – a member of many more local communities. Korsgaard calls personal relationships "a Kingdom of Two" and adds a wonderfully honest remark:

> "And the thought of oneself as a certain person's friend or lover or parent or child can be a particularly deep form of practical identity. There is no obvious reason why your relationship to humanity at large should always matter more to you than your relationship to some particular person; no general reason why the laws of the Kingdom of Ends should have more force than the laws of a Kingdom of Two. I believe that this is why personal relationships can be the source of some particularly intractable conflicts with morality" (Korsgaard 1996: 128).

Even if these practical identities which are fundamentally at odds with the human identity must be given up[106] (for example, an identity of a mafioso), there are those practical identities which do not contradict the value of humanity, but which come into conflict with it under particular circumstances. At this point, though, one can get confused: if all rational actions may be moral because of their form, it means that actions which spring from particular identities may also be moral in virtue of their form, but if so, then what does this moral identity amount to? One can have doubts if it is needed at all. After

[106] I have to mention still another principle of practical reason, the need of which Korsgaard admits. This principle would be a response to the acknowledged need for balancing our various reasons for action, or for ranking our various practical identities. Korsgaard believes that all the normative principles of reason have to be formal and not substantive, therefore she rejects all of the formulations of the principle of prudence so far given. But she acknowledges the need of its purely formal formulation, because it is one of the principles that is necessary for the very existence of agency: "we characteristically have more than one aim, and ... rationality requires us to take this into account when we deliberate" (Korsgaard 2008h: 29) and "a formal principle for balancing our various ends and reasons must be a principle for unifying our agency, since that is so exactly why we need it: so that we are not always tripping over ourselves when we pursue our various projects, so that our agency is not incoherent" (Korsgaard 2009: 58). However, I find a more developed analysis of this principle unnecessary for our aims in the present work.

all, an identity gives substance to our judgements, it is the source of reasons the source of the strength of normativity of those respective reasons, so a moral identity seems to be, on the one hand, superfluous, on the other hand, somewhat substantial.

The moral law gives the intersubjective dimension to the decisions, makes them justifiable in principle. It also introduces the other-regarding aspect into the process of decision making. By a universalization procedure represented by the categorical imperative "you ask whether you could will to be part of an order of things in which this was the universal practice, and at the same time rationally will the maxim in question yourself" (Korsgaard 2008a: 222). In other words, you ask yourself if you wanted others to treat you in the same way that you treat those others in virtue of everyone acting on the same principle. The moral law brings another consideration in: would all rational people agree to be part of such an order of things, would others want to be treated that way. So why being, for example, an intersubjectively good mother is not enough for being a morally good person? Surely, the various actions of mothers have a moral dimension.

In other words, why not think that the moral identity as such does not exist, that what makes a judgement moral is always just a form, but never the contents? The more so, as Smith and Korsgaard include specific circumstance of a deliberator into the maxim that passes the universalization test. However, I am afraid that Korsgaard does not make explicit (or I did not manage to restore this explication) what the moral identity exactly amounts to. In a conversation, asked if other practical identities are also sources of moral reasons, she answered to the negative. So I tend to think that one should rather think of the moral identity and its reasons as a certain standard that one can test oneself against. If you are thinking what to do and you have a reason to choose something that every good mother would, you are "locally" justified, but you can also ask yourself if this thing is also something that a morally good person would choose, if you are to be justified also "more globally". If there is a gap between the decisions, is there is space for remorse, reproach from one's own

conscience, then the choice is not moral in the universal sense. When making a moral decision we become as stripped from our contingent identities as other humans and it is in relation to such other human beings who are not religiously or otherwise committed, politically or otherwise biased, that the global justification may take place.

There are some reasons why this moral identity is needed. First, it is needed as a source of purely moral incentives – otherwise, where would they come from? Second, identities are expressions of our relational definitions, and our relation to others as simply human beings defines us as well – even if not as often as other relations. Third, an intersubjective character of judgements sprung from particular identities may be not enough for the impartiality which defines (universalist) morality. After all, Kant was criticised on this account and his adherents were trying to add other requirements which would guarantee the right kind of intersubjectivity for the moral maxims. As Blackburn said, "if everyone came to think that it was permissible to maltreat animals, it would not stop being bad – it would only mean that everybody has deteriorated" (Blackburn 1987: 14). Or if all parents thought that it was right to favour their own children when selecting a candidate for a job (just in virtue of them being their children), it would not become morally right. So on my understanding, the moral identity represents that part of us which is stripped of the particular identities, which defines us a human among other humans and which embodies the source of maximally impartial other-regarding considerations.

Surely, one may still feel it more important to be a good mother in the eyes of one's own society, family and even one's self than to be a good human being in a more general sense. Then one enforces one's practical identity of a mother and takes responsibility for one's choice. As Korsgaard said, it is not obvious why your relation to humanity should be more important to you than your particular relation. The human identity should probably serve as a litmus paper to the universal acceptability of our choices. This theory, thus, by leaving the possibility of a gap between the normative judgements with

different degree of justifiability leaves space for rational criticism of one's locally justified choices, space for moral growth and change.

At the same time I believe that we do have this moral practical identity and that is why it belongs in a good theory. Even if not often, we evaluate ourselves as human beings, care about more than our everyday chores: e.g. I am a good technician, a good friend, a good Lithuanian, but am I a good person? What is my purpose in life? What matters the most? What is a human for? Moral questions are among the existential questions – as morality is, after all, one of the defining features of humans. Also, I think that we sometimes feel just as humans among other humans, and view others just as such – especially faced with their needy presence. And it is the existence of this identity that enables any reflective agent to be "led to acknowledge that she has moral obligations" and shows how morality is special: "it springs from a form of identity which cannot be rejected unless we are prepared to reject practical normativity, or the existence of practical reasons, altogether" (Korsgaard 1996: 125). That is, the existence of an independent moral identity makes existence of moral reasons independent on one's having any of the particular practical identities; it makes the existence of moral reasons necessary in so far as every human animal necessarily has a human identity.

Conclusions

1. Moral realism, understood as a methodological position, embodies an aspiration to coherence between the actual functioning of morality and our knowledge of this functioning. The idea of coherence manifests itself as social and intrapersonal transparency, or as harmony of beliefs and practices both on the social and on the personal levels. Therefore, a theory resulting from a methodology which presupposes the value of coherence is preferable to theories that use other methodological approaches, as the latter are doomed or highly likely to leave a split between the beliefs and practices.

2. While analysing the two possible versions of moral realism, it is found that the one defending a view that the truth making conditions of moral judgements are mind-independent (MR_{MI}) has several flaws, only one of which is that it cannot incorporate the action-guidingness of moral judgements.

2.1. In virtue of its naturalistic construal of objectivity, MR_{MI} cannot answer the normative question ("why to be moral?"). Thus, the authority of morality becomes conditional either on every person's extra-moral interests (then morality has no authority of its own), or on every person's moral dispositions or desires (then moral agents are turned into moral fetishists who care to do whatever is right). This means that MR_{MI} alienates the moral sphere from the world of direct human concern and suggests an unacceptable psychological picture of moral agents.

2.2. If one takes into account the claim of MR_{MI} that how one conceives of a normative question varies with one's metaphysics and so that for MR_{MI} the normativity of morality is justified metaphysically, one has to evaluate the plausibility of its metaphysics in its own right. However, it appears that the moral ontology cannot justify normativity, because moral ontology itself hangs on faith: there is no plausible epistemology available that would connect us to such a moral ontology; there is no positive picture of such an ontology available, but only several unsuccessful analogies between the

moral and the natural world; the semantic theories that fit terms for natural kinds and properties do not fit the moral terms; finally, moral ontology is not necessary for explaining the truth-aptness of moral judgements.

2.3. MR_{MI} cannot account for the apparent practicality of moral judgements given even the third kind of criteria: it does not pass the Open Question Argument interpreted as a test of normativity for descriptivist accounts. Thereby MR_{MI} fails to accommodate the second feature of common-sense morality, i.e. the action-guidingness of moral judgements.

3. The mind-dependent version of moral realism (MR_{MD}) does not share the aforementioned disadvantages. This position manages to harmonise our common-sense supposition about the action-guiding character of moral judgements with its theoretical view of moral reality. MR_{MD} separates ontology from ethics, and it is this conception of moral reality as dependent on the human mind in a non-relative manner that supplies plausible explanations and justifications of the issues that the MR_{MI} could not account for. MR_{MD} explains the mechanisms of instantiation of moral properties without relying on murky analogies, gives a clear explication of how moral judgements can have truth values and how we know in particular instances whether they are true or false, preserves the normativity of morality and provides a plausible picture of the psychology of morally good people.

4. MR_{MD} and MR_{MI} differ as to the question of what epistemic pathways lead us to knowledge of the moral reality: one relies on reason, the other prefers senses. Analysis of the causes of the plausibility of MR_{MD} and the failures of MR_{MI} (in relation to the construal of the truth-aptness of moral judgements and to an inclusion of the element of practicality of moral judgements into theory) brings us to the conclusion that it is this difference in epistemic approach that is crucial. Thus, rationalist epistemology is preferable in the moral sphere.

5. In order to explore how MR_{MD} manages to account for the practicality of moral judgements, one has to investigate the viability of motivational internalism (MI), a position defending the necessarily motivating character of

moral (and more widely – practical) judgements. Analysis of the varieties of and counter-arguments to MI has showed that:

5.1. A restricted version of moral MI is needed in order to avoid an oversimplified picture of human psychology and to supply the MI accounts with explanatory power. A proportionality/commensurateness requirement, requiring that the strength of the evaluative practical judgement has necessarily to be proportionate to or commensurate with the strength of the motivation to act accordingly, amends the aforementioned defects.

5.2. A conditional version of MI (the necessary relation between moral judgement and relevant motivation holds under a certain condition) is needed. Unconditional moral MI should be supplemented with a proviso to accommodate the cases of moral indifference.

6. A restricted variant of rationalist internalism (RI), the most promising version of conditional MI, is built on the central notion of rationality. Rationality should be identified *primarily* with psychological coherence and not with normal mental functioning; accordingly, irrationality does not amount to mental abnormality, but rather only to incoherence, the falling short of the requirements of coherence. With respect to what elements should cohere, we can analytically discern different kinds of rationality requirements or principles. Full rationality in the practical context boils down to pervasive coherence (synchronic and diachronic intrapersonal, interpersonal coherence), i.e. coherence between all of the relevant mental states involved in a certain practical decision.

7. The conception of rationality is to be supplemented with the element of autonomy/orthonomy, as coherence is not sufficient for rationality. Rationality, after all, involves an element of human activity as it is an attribute of the workings of reason, of the active faculty of the human mind. Autonomy/ orthonomy, as self-rule by appropriate principles, supplies the needed element and gives more substance to the relation of coherence.

8. According to RI, we are special animals, humans, in virtue of our possibilities and (reflective) structure of mind which enables control over ourselves. Such a mind is the source of rationality. As moral value pertains only to human actions, morality is thought to have the same source as rationality – reason. To a certain extent rational actions and moral actions are defined by the same principles of reason, e.g. by the categorical and hypothetical imperatives which grant a person unity and efficiency. The specifically moral judgements, though, are marked by distinctly universalist and other-regarding character. In relation to rationality, morality is defined by further requirements of coherence (adding the interpersonal kind), so a moral action is a perfect rational action, a human action *par excellence*. And morality is the supreme form of rationality.

Given the conclusions (1) to (8), I conclude that a moral theory which embodies the two fundamental features of common-sense morality (truth-aptness and action-guidingness of moral judgements) is possible and that it is rationalist internalism which manages to combine them in virtue of making coherence its constitutive value.

Appendix 1

Meta-ethics: conception and wider methodological context of the present research

Meta-ethics: primary tasks, relation to common practice and to normative ethics. Meta-ethics can be defined as "the attempt to understand the metaphysical, epistemological, semantic, and psychological, presuppositions and commitments of moral thought, talk, and practice" (Sayre-McCord 2012). In other words, it is an activity of "reflecting on the presuppositions and commitments of those engaging in moral thought, talk, and practice and so abstracting away from particular moral judgments" (ibid.). And, I would emphasise, that, regardless of what the "metaethics's substantive assumptions and practical implications might be" (ibid.). Thus, even if meta-ethics as a distinct and recognised branch of ethics was born at the beginning of the XXth century (the so-called father of meta-ethics being G. E. Moore with his *Principia Ethica*, 1903), as an enterprise of trying out the foundations of morality, or ethics[107], it cannot be thought to be a prerogative of philosophers or just of interest to contemporary people.

First, foundations and respective pretensions of morality, or ethics, were being challenged or just discussed already in the works of Plato by such figures as Trasymachos, Glaukon, Kalikles, (naturally) Socrates and others. Needless to say, meta-ethical considerations came up in the works of many other Western philosophers writing well before the XXth century, the great ethicists being also meta-ethicists (Aristotle, Kant, Hume – to mention just a few of the most outstanding names).

Second, meta-ethical considerations are not alien to non-philosophers' lives: from time to time objectivity of morality, its source of normativity and

[107] Here I mention both morality and ethics because some theorists, such as Bernard Williams and others, thereby express a distinction between the subject matter of ancient and modern philosophy concerning questions of proper behaviour of human beings. And, accordingly, the subject matter of the thoughts of the folk in that respect.

other questions of the kind come up in people's conversations – implicitly or openly (probably not in such technical terms, though). As Sayre-McCord notes, "its central concerns arise naturally — perhaps even inevitably — as one reflects critically on one's own moral convictions" (Sayre-McCord 2012), or, one could add, as one encounters those who think differently about moral matters. Bernard Williams incisively notices that not all reflection requires ethical theory. Some kinds of reflection may require explanation which can be given in social or psychological terms, so it is only that type of critical reflection which seeks justification that leads to ethical theory (Williams 2006/1985: 112).

Meta-ethical considerations are called for, to put it more metaphorically, when the faith wavers, when we step back from our usual practices, when the one who challenges your views comes up to you either from within you or from without. While taking this step people enter the sphere of the theoretical inquiry about practical matters, or, if you wish, the sphere of meta-considerations – with the tools available to them. Supposedly, philosophers are better equipped to deal with theoretical problems, but it does not mean they are solving different problems than those that arise for non-philosophers or to philosophers-after-work. As Michael Smith claims, philosophers even do not give different answers to moral questions than ordinary folk does, but those answers are more technical and systematic. He insists that "The task of the philosopher in meta-ethics is to *make sense* of ordinary moral practice" (Smith 1994: 11; emphasis mine).

Therefore, the primary task of meta-ethics is that of making sense of the moral practice that exists and persists. Certainly, philosophers do not take up to explain all the aspects of that practice (or practices), so I will need to point out which aspects seem to matter most and why. Also, one should note that meta-ethics as a discipline consists of a great variety of views, so one can understand that this vision of meta-ethics at some point breaks down to more specific tasks and different theorists differ in which of them are most important – or accordingly, which aspects of the practice are the most important to

explicate. Surely, as a consequence, the explanations of the moral practices by different philosophers, as well as by different persons from the folk, need not be the same. However, theorists (surprisingly or not) converge on which two main aspects of our moral practices need to be explained in moral terms or at least explained away in non-moral terms.

To claim that such is the task of meta-ethics, is not to say that each and every theorist begins from identifying the most common practices, uncovers the premises and tries to put them into a coherent theory. It is apparent that philosophy is an on-going debate where one never begins from nothing, so this task is rather an underlying and most general aim that meta-ethicists have; it is its genealogical scheme. An individual theorist may begin from defending a position, not the most common to everyday moral thought, such as scepticism of some kind or a radical form of expressivism.

Also, to say that the primary task of meta-ethics is to understand, reflect on the presuppositions of moral language, thought and practice, does not mean that it is purely descriptive. Depending on the results of the inquiry (for example, after having argued for a sceptical position), different theorists may still make differing recommendations as to our further moral practices given the truthfulness, falsity or a shaky character of their premises. Not all theorists, for example, after concluding that we err in grounding morality on certain premises, require us to change those mistaken premises (S. Blackburn thinks we should retain them for practical purposes – as if they were true) or to change our common practices (e.g. J. L. Mackie thinks we should retain them). But those who seek transparency between our beliefs and practices, will surely feel obliged to change practices in accordance with the moral truth which they believe to have found out.

So as for the implications of meta-ethical findings to normative theories and to common moral practices, they are disputable. In quite some cases meta-ethical answers are compatible with the possibility of more than one normative theory, though it is, of course, possible for them to exclude some (other) normative theories. And whether a theory can change the practice itself in

general, depends both on a theory and on the nature of practice. (These are very general and vague considerations, but I do not wish here to go into details.)

However, the relation of meta-theory and normative theory can be even more complicated, for example, Smith believes that normative ethics is "crucial for the final resolution of meta-ethical questions" (Smith 1994: 3). In his case it means that his meta-theory of morality offers a theoretically neat picture of what rightness is like given our common practices, however, if such rightness is not only theoretically possible, but also can be substantiated in our actual world, depends on whether the proposed analysis of rightness can be filled out (for example, if there really are reasons that all of us share, and so on). In other words, meta-ethical enterprise is often (and will be in this research) just conceptual, meaning that meta-ethicists seek *to make theoretical sense of suppositions*, however, it is an open question if they have referents[108] actually (we can only present some inductive arguments in support of such hopes).

Methodological approaches. Before proceeding to the more specific considerations, I will consider a more general problem – that of how to best approach reality (in the widest sense of the word) with the task of building an adequate theory, or, if you will, of discerning the cases of genuine knowledge from sophistry. This discussion serves the function of placing the rationalist internalism (RI) into a certain tradition of philosophising and that of justifying its approach, i.e. defending its right to make its claims (and so it should keep part of the possible criticisms at bay). At the same time, it is an opportunity to bring forward my own reasons for the choice of such a structure of the present research.

[108] That is, if something is conceptually possible, it is not necessarily ontologically possible in our world. Surely, if something is possible in our world, it does not mean it is also actual, but there being no actuality is not as crucial as the impossibility for undermining the relevance of a moral theory.

As meta-ethicists are primarily concerned with the status of the metaphysical, epistemological, psychological and semantic presuppositions of moral practices, they are interested in proposing one or another theory of moral reality[109] and of its relation to moral language. In other words, they are to sort out the knowledge of morality. In general it is possible to approach the question of how to build (or test) a body of knowledge, or to construct a theory yielding knowledge about the world (moral or otherwise) that we live in in several different ways. Roderick Chisholm, first of all, discerns three such logical possibilities which he discusses in relation to what he calls "the problem of the criterion". In his "Theory of Knowledge" (1977/1966), in Chapter 7 called "The Problem of the Criterion", Chisholm discusses the features common to questions about knowledge. He also tackles the same problem in his Aquinas Lecture with the same title (2001/1973), but turns to slightly different considerations. I will refer to both works as they complement each other.

He puts the two most general questions of epistemology like this: "*What* do we know?" and "How are we to decide, in any particular case, *whether* we know?", or, to put it otherwise: "What is the *extent* of our knowledge?" and "What are the *criteria* of knowing?" (Chisholm 1977: 120). Chisholm claims that in order to answer one of these questions one is required to answer the other one, so one is necessarily caught in a vicious circle: in order to know if things are really the way they seem to be, we must have a procedure for distinguishing the true appearances from the false ones, but in order to know if our procedure is good, if it succeeds in distinguishing them, we should know which appearances are true and which false (Chisholm 2001: 190). One of the possible views to this problem is scepticism which takes the gravity of the problem to block the possibility of any solid solution. However, this "is only

[109] No particular metaphysical view should be presupposed here. It may be that the moral knowledge and moral reality will be shown to be impossible and so will be just explained away.

one of the three possibilities and in itself has no more to recommend it than the others do" (ibid. 198).

Another possibility, called by Chisholm particularism, is to answer the first question of the extent of knowledge and, based on that, to answer the second one (of the criteria of knowledge); the third possibility, called methodism, is to begin with the second question and proceed to the first. Chisholm himself favours the particularist start and criticises empiricists as an instance of a methodist approach. His reason for this is that empiricists, applying their arbitrary and very broad criterion, have drastically reduced the extent of our knowledge of the world, thus, they have excluded from the view of the world part of the putatively real appearances. The risk that the methodist approach involves is that its resulting theories can have little to do with reality: perhaps they will circumscribe as reality too narrow areas or even will not pick out any reality at all. I believe that the latter problem is less likely in the sphere of knowledge about the natural world than in the sphere of ethics (for example, empirical criterion is likely to result in at least some natural knowledge, but not necessarily in ethical knowledge – as in case of expressivism).

In this light, the safer way is to start from particularism and then to proceed to the question of epistemic pathways. When answering the second question, according to Chisholm, it is possible to choose from the four traditional sources of knowledge (external perception, memory, self-awareness/reflection and reason) or to declare existence of a new one – in sight of one's acknowledged reality (say, intuition). One can disagree with Chisholm about the list of the sources of knowledge (or to emphasise that reality can be exhaustively, and thus properly known only combining some or all of those sources), but it is not what is the most important for me here. Particularism, or common-sensism, in ethics consists in acknowledging that we do know certain ethical facts (or in acknowledging some moral phenomena the status of reality based on common sense).

In ethics, again, three sharply different positions are possible logically[110]. Chisholm presents them in the following sequence, which, one can note, repeats the sequence of and reveals the logic behind the structure of the historical development of meta-ethical debates; it is obtained using "the G. E. Moore shift"[111]:

"The 'intuitionist' will reason in essentially the following way:

(P) We have knowledge of certain ethical facts.

(Q) Experience and reason do not yield such knowledge.

(R) Therefore, there is an additional source of knowledge.

The 'skeptic', finding no such additional source of knowledge, reasons with equal cogency in the following way:

(Not-R) There is no source of knowledge other than experience and reason.

(Q) Experience and reason do not yield any knowledge of ethical facts.

(Not-P) Therefore we do not have knowledge of any ethical facts

… one could also argue in this way:

(Not-R) There is no source of knowledge other than experience and reason.

(P) We have knowledge of certain ethical facts.

(Not-Q) Therefore experience and reason yield knowledge of ethical facts" (Chisholm 1977: 124).

The first and the third positions can be called realist in the widest sense of the term (as acknowledging that ethical knowledge is available to us) with its roots in common sense (certain moral truths that we hold are actually true). The third argument (and here I depart from Chisholm's text) can be realised in

[110] If we acknowledge that ethical facts can be known either through experience and reason or through "an additional source of knowledge". We exclude from the range of possibilities such positions that do not think ethical knowledge (whether in a looser or more robust sense) is to be had at all (such as expressivism, for example), but only because they are not instances of particularism in the first place.

[111] It is a term coined by William Rowe. The shift consists in rearranging the propositions in two arguments so that one of the premises of the previous argument is retained while its conclusion negated thus obtaining a new conclusion which is a negated other premise of that previous argument.

different ways: one can choose not only one of the so-called traditional sources of knowledge instead of two or a different combination of them, but also one can interpret "certain ethical facts" differently. For the sake of classification, we can discern the logical possibilities the way Chisholm did, but actually things are rarely so clear cut. Say, John Rawls's method of reflective equilibrium is a well-known example that "transcends the traditional two tiered approach to moral inquiry according to which one must choose as one's starting points either particular moral judgments or general moral principles" (DePaul 1988: 67). Thus, the meta-philosophical position of common-sensism can be realised in a theory – or a meta-theory for that matter[112] – in several ways.

However, I am well aware that "moral realism" is a problematic label. Nowadays, it can be attached to positions ranging from Moorean robust moral realism often associated with Platonism to those moderate ones which are simply adverse to relativism. However, when I used this term in its widest sense, I meant it as a methodological position. In this sense realism is held to be a position beginning the quest for knowledge from assigning some of the phenomena the status of reality or verisimilitude. Perhaps it is better to give it a different name. It can be called common-sensism due to the fact that this position gives credit to a common-sense view of the world, or in virtue of the importance it bestows on common sense at the beginning of the theoretical quest. I think, though, that both names are appropriate. And it is a matter of further inquiry what epistemic pathways lead to such knowledge.

These methodological options being set forth, I want to bring out one aspect that this classification draws attention to. One can note that, according to Chisholm, the choice of a starting point for an inquiry (from the three

[112] I want to underline this whole claim, as well as its part where I say that this methodology can be applied regardless of whether we talk about particular moral judgements and general moral principles, like, for example, that „Killing innocent people is wrong" and „That which maximizes happiness of the humanity, is right/moral" (then, we are interested in building a moral theory), or about particular features of moral judgements and general principles defining moral judgements (then, we are interested in building a meta-theory).

options classified as scepticism, methodism and particularism, as mentioned) is arbitrary. As DePaul puts it, "there is no way of proving which of these positions is correct. When a proponent of any attempts to do so, he or she will eventually be forced to beg the question against the others" (DePaul 1988: 68, n. 7). He points out that there is no non-question begging reason to favour one starting point over the other (that is, once you question your opponent's position, you assume one of the other two positions), so "The decision of what to take as one's starting points for epistemological inquiry is ultimately, in an important sense, arbitrary, but since there is no avoiding this arbitrariness, one should choose wisely and forge ahead" (DePaul 2009: 44).

This works as much in favour of my own project as against it. On the one hand, it means that common-sensism is (at least) not worse off than any other alternative view. On the other hand, such a Chisholmian view does not allow any meaningful communication between these basic positions: it seems impossible to give reasons in favour of common-sensism that would convince a sceptic or a methodist. Perhaps one can judge the three strategies without begging the question according to some external criteria, such as criteria of a good theory or so? That may be the case, but I will not set myself a goal of tackling the pros and cons of the three in detail, but will still put forth some of the advantages of common-sensism and will shift the burden of disproof to the other two positions, holding that until proved otherwise, common-sensism about morality in not only on the same footing as two other positions, but superior to these – given the already presented and the forthcoming considerations.

One of the adherents to common-sensism, Thomas Reid thought that showing that some sceptical hypothesis is on the same footing as the common-sense belief, is enough to justify the latter (Yaffe and Nichols 2009). He believed that justification did not "necessarily require providing positive reasons in favor of common-sensical beliefs; common sense beliefs could be adequately justified simply by undermining the force of the reasons in favor of alternatives to common sense" (ibid.).

Bernard Williams, however, in his (2006/1985), considering the question of the good and bad beginning of ethical theory, has proposed more explicit criticisms to the strategies of beginning from within ethical theory and from without it, and finally turned to "ethical experience" as a starting point.

In addition to the belief that such a strategy of inquiry is superior as more probably leading to a true theory, there may also be other considerations that speak in its favour. These are more or less practical.

One of the often heard motivations for choosing a moderate position in practical philosophy is the fear for the possibly catastrophical practical consequences of a theory. As Sayre-McCord puts it, if the presuppositions and commitments that people take for granted turn out to be suspect, then not only their understanding of that part of their lives is compromised, but also the sense of its importance may disappear as well (Sayre-McCord 2012).

It does not mean that such fears should override theorists' commitment to truth or that practical philosophy should be directed at preserving the *status quo* of moral or political practices. It only indicates reluctance to experiment on living beings and frail social systems. And an intuition shared by many is that in practical philosophy being a conservative is indeed more justified than in epistemology (whether because the nature of the non-human reality is less versatile with regard to theoretical manipulations or because of other reasons). The greater reason for such fears have those who sympathise with the constructivists, as the latter see ethics as "performative" rather than descriptive. Ethics for them functions *creating* the reality, whereas the non-moral theories end up just either falsely or correctly describing reality. Thus, one should have good reasons for a change, such as, for example, a change of the suppositions that our ordinary practices rely on (and which we will talk about shortly), that is, a change that would *affect* our ordinary moral practices.

True, some of the theorists, as already mentioned, do not propose to change the practice in view of the counter-evidence to the faulty bases of the practices. Though what it does is bringing about dissonance in people. However, a theory that could preserve our integrity – both as theorists (and

here I do not just have philosophers in mind, but also the folk) and practitioners – would have an advantage.

Aspiration to maximum coherence as an advantage of common-sensism. As for scepticism, everyone knows that it does not yield a viable ethical (not meta-ethical) position: one cannot suspend one's judgement in face of one's own need to act, and one's omitting an act is in most cases as subject to ethical evaluation as committing an act. So scepticism puts a person into a strange position, or a strange state of mind, of acting on what very well may be or even is a mistaken knowledge. I call it "schizophrenia" in its etymological sense of "split mind": a person believes one thing, but acts on another, and – what is more – by one's own lights. Certainly, one may say that a person may be sceptical only with regard to reason as the source of morality and with regard to the possibility of moral knowledge thinking that ethics is based on emotions or gut feeling. However, the "schizophrenia" in such people's minds arises in the context of our common practices and moral phenomenology (people from time to time ask for reasons and believe that there are intersubjectively or even objectively correct and incorrect answers to be had to moral questions etc.).

Meanwhile, the common-sense beginning embodies the aspiration to transparency so that the actual functioning of morality and our knowledge of its functioning correspond to each other from our own point of view.

Harmonisation of the beliefs, or coherence[113] of beliefs and practices, is closely connected to effective and autonomous acting (which, for some

[113] When thinking of coherence, it should be conceived of widely. Here Harman's considerations may be helpful (regardless of his talking about coherence of beliefs and intentions, we can understand that the elements which cohere can also be other states of mind or their contents): "we can … distinguish two sorts of coherence, positive and negative. … Negative coherence is merely the absence of incoherence. Beliefs and intentions are incoherent to the extent that they are inconsistent with each other or clash in other ways. … Positive coherence among one's beliefs and intentions exists to the extent that they are connected in ways that allow them to support each other. Relevant connections may involve explanations, generalizations, and implications" (Harman 2002: 180).

philosophers, is a platitude as behaviour that lacks these qualities is not acting at all). Theorists in pursuit of harmonising moral theory with moral practice try to preserve the inner coherence and autonomy of the moral agents. Incoherence is inhibiting, invalidating agency. According to Korsgaard, you have to be whole or one in order to act at all. To clarify what she means, she refers to Plato's *Republic* where the human soul is compared to the constitution of a *polis*. One – whether it is a city-state or a person – is able to act only if one's behaviour springs from one's constitution. "A constitution defines a set of roles and offices that together constitute a procedure for deliberative action, saying who shall perform each step and how it shall be done" (Korsgaard 2004g: 105). If every part of the soul or a city-state sticks to its function, it will act as a single agent, it will be effective. Otherwise, there will only be inner conflicts to no effect or to no benefit to the whole which will threaten to destabilise it.

Similar is the idea behind Smith's thinking. For him, self-ruling is necessary for autonomy, but not enough, what is needed, is also sticking to one's function in bringing about the deliberations obtained according to the right procedures: "Our image of non-heteronomy is driven by a more traditional metaphor of good government than the democratic metaphor which seems to inspire such visions. The good government of desire is a regime under which desire is faithful to the rule of deliberation; being endogenously inspired and maintained is not enough, even if it is necessary" (Pettit and Smith 1993: 76).

And to the contrary, theories that separate the truth of the theory from the truth of the practice, threaten the effectiveness or even autonomy of the agents and make ethics a subject of political agenda (what behaviour it is best that people stick to?) or a subject of science. Bernard Williams calls this value transparency that is undermined by the "schizophrenic" theories. According to Bernard Williams, it is primarily social transparency: "the working of its ethical institutions should not depend on members of the community misunderstanding how they work" (Williams 2006: 101). Here he has contractualism, an ethical theory, in mind, but this fits talking about the

motivation behind higher order theorising as well. Some theories, like sorts of utilitarianism (e.g. Sidgwick's), lack social transparency by dividing people into "theorists who could responsibly handle the utilitarian justification of nonutilitarian dispositions, the other a class who unreflectively deployed those dispositions" (Williams 2006: 108). So the truth about morality should be kept from the public allowing the class of theorists to live by different rules and to make exceptions for themselves, but not for the rest of the people.

Another part of utilitarianism (like Hare's), as Williams claims, lack psychological transparency when the gap is made not in social terms, but in psychological ones. That is, the theory may differ from the practice *within* the same person (it is an intrapersonal incoherence rather than interpersonal incoherence): there is "the *time* for theorizing and the *time* of practice", there is a cool hour when the agent "leaves himself and sees everything, including his own dispositions, from the point of view of the universe and then, returning, takes up practical life" (Williams 2006: 109 and 110). However, Williams thinks this is an artificial barrier, a surrogate to the class barriers of Sidgwickian theory, an illusory dissociation, because actually process of theorising is a particular kind of practice (ibid.).

The ones who base their behaviour on such a utilitarian theory do something to the contrary of what the theorists in pursuit of harmonisation think an agent has to behave in accordance with, i.e. such people do not behave in accordance with the principle they identify with and which provides standard for a good action. Instead, they are either deceived by somebody else (and manipulated) or by themselves and so act in accordance with what they are made to believe is right, with whatever regulations they are given (or even in accordance to what they know to be not right, or unjustified).

Williams calls the first of those utilitarian positions "Government House utilitarianism" connecting it with the colonialist connections of this position. It implies a very different analogy from that of a constitutional democracy or a well-ordered city-state. The second utilitarian position is closer to the ideal of ethics as science dislocating the theorising self (if that can be called a "self")

beyond the practicing self, clearly separating the theoretical enterprises from the practical ones. The lack of transparency (with regard to the true bases of one's behaviour) for the agent in such theories, or the lack of coherence within an acting unit (of whichever size), then, is not seen by utilitarians as a major disadvantage, and it certainly modifies their understanding of agency: unity, integrity or inner coherence is not necessary for agency.

Such a split is rather likely to be obtained also by methodist approach theories, because they are likely to produce a very restricted and in some aspects counterintuitive view of reality which conflicts with some of the fundamental aspects of common-sense understanding of reality. This kind of split can either discredit the philosophical conclusions or issue into an intrapersonal conflict of such split theorists. For example, Moore, discussing the sphere of theoretical philosophy, notices a certain inconsistency or even outright self-contradiction in the views of those philosophers who deny the most basic and evident common-sense truths. He claims that frequently some of those philosophers held "as part of their philosophical creed, propositions inconsistent with what they themselves *knew* to be true" (Moore 2002: 41). And others still, according to him, are contradicting themselves by uttering with certainty general presuppositions which are supposed to deny the possibility of certain non-egoistic knowledge (e.g. "There have existed many human beings beside myself, and none of us has ever known of the existence of any human beings beside himself").

The thing that Moore emphasises is that he is not alone to think that "in certain fundamental features" common-sense view of the world is wholly true, because all philosophers do. However, the difference between him and many others supposedly lies in the fact that those others hold *as philosophers*, in addition, views inconsistent with those same fundamental features. Now, one should be fair and acknowledge that contradictions between one's common-sense view of the world and some other, theoretically (in)formed, view of the world are inevitable. And surely, Moore himself acknowledges that not all common-sense beliefs are true. The question, then, can be reformulated as

which of the contradictions (or intrapersonal schisms) are acceptable and which are not. Moore's answer would be that the contradictions of *fundamental* features would be unacceptable. But which features are fundamental? Moore gives a list of examples in "A Defence of Common Sense" and, based on it, a general claim, but not a list of fundamental *moral* features. So what could be the criterion or criteria of this fundamentality at hand? This is the first and the most important question which we will address shortly.

Second, what is the relation of the common-sense world view, a common-sense theory and practice? That is, selecting only some (fundamental) features of reality to put into theory, common-sensists already do something to correct that primary naïve worldview. So the product of their work, the resulting worldview, should differ from completely common-sense worldview, as their work consists in bringing about that difference, or that change. Common-sense knowledge differs from theoretical knowledge and adherents of common-sensism have work to do to shift from one to another. As a result of the introduction of such a theory, some of the common-sense beliefs may be rejected or the resulting theory may be proved to be wrong, e.g. not working practically perhaps because of taking the wrong phenomena as real. Or perhaps both of the views would be retained by the people?

For example, Moore thinks of the work of a philosopher as that of giving correct analyses of the widely understood expressions. He holds that some of the expressions (such as "The earth has existed for many years past") are unambiguous and that we all understand the meaning of them in so far as they are used in their ordinary (and not, say, some specific philosophical) sense. The meaning, though, should not be confused with the correct analysis of such expressions. The latter is already a philosophical task. However, Moore claims that we cannot raise the question of analysis if we do not understand the meaning in the first place, thus, the meaning comes first. Moore being a realist in both theoretical and practical philosophy, his views on the task of a philosopher apply accordingly to both.

Moore's methodological realist's program fits different versions of methodological realism about morality – regardless of the possible difference in their views about the functions of the various concepts which is not of importance here. For example, Korsgaard, a constructivist (a kind of a methodological realism) in practical philosophy, thinks that the function of at least some of the practical concepts is not to describe the world, but to mark out the solution to some problems that people face (Korsgaard 2008e: 22). And a philosopher's task is to "identify the content of such a concept by working out the solution to the problem, thus providing a particular *conception* of whatever the concept names" (ibid.).

Another part of a philosopher's work is to build such a theory that would weed out the possible inconsistencies of a common-sense, or folk theory of the world, also, to fill the gaps of it and to provide discursive justification to what it only intuitively holds. I should also remind that such method is neither new nor especially controversial. In this context one cannot fail to remember Socrates/Plato and Aristotle. We know, for example, that Aristotle was favourably using the term *endoxa* (sing. Gr. ἐνδόξων) referring to the commonly held beliefs, or opinions[114], which served as the starting point for dialectical argument. Beliefs of the people or of the wise, or the most reputable, though, can notoriously be conflicting or wrong, nevertheless, possibly not all of them and not completely. And that is where the philosopher can do one's part. Like Moore has put it, to find the right analyses, the justified beliefs.

It seems that in some of Plato's dialogues Socrates does a similar thing: the ones he converses with seem to know how to use one or another concept, but giving definition of it does not come so easily, because first, many prejudice and inessential elements of definition must be shed. And in the late works of Plato, it is only after collecting all the relevant instances of some

[114] "Generally accepted opinions, on the other hand, are those which commend themselves to all or to the majority or to the wise – that is, to all of the wise or to the majority or to the most famous and distinguished of them" (Aristotle 1960: 273-275, that is, an English translation of the lines 100b21–23 from *Topica*).

category and after the analytical procedures that the definitions can be obtained. Thus, again the task of a philosopher is to purify the concept, removing from it the inessential elements of meaning and the contradictions that sometimes contaminate folk conceptions.

Common-sensism is an optimistic, or even a naïve position, which trusts that our moral practices basically are on the right track, that people can discern the main aspects of moral reality and so that in their main beliefs (as to the character of morality) they do not err. It allows people to have access to that reality without any specific tools, without being privileged. Thus, it purports to give a transparent theory, i.e. the one where the nature and requirements of morality are accessible to the ones subject to it, preserving the integrity and autonomy of the moral agents.

On the one hand, of course, a theoretical inquiry may show those pre-theoretical beliefs to be incompatible and issue in our shedding some of the common-sense beliefs, and a subsequent moral theory may become in some respects counter-intuitive. Or the world may finally prove us to be wrong in believing something which supposedly presents us with conclusive proof. But till that is done, we are justified in believing some things more than others, or at least, we are not less justified to. I find compelling the view of DePaul: "We therefore do well to avoid adopting … pessimistic positions … before we've given the more positive alternatives a fair shot at completing their projects" (DePaul 1988: 73-4).

Appendix 2

Survey by Bourget and Chalmers

Survey and the aim of its use, restrictions. In 2009 David Bourget and David J. Chalmers have conducted a survey on philosophical views of the various analytical philosophers, which was followed by a meta-survey (what did philosophers believe other philosophers believed?). In 2013 they have prepared a draft of paper entitled "What do philosophers believe?" which sums up the results and discusses their meaning, restrictions and importance. Three questions from the questionnaire concern meta-ethics in particular: "Meta-ethics: moral realism or moral anti-realism?", "Moral judgment: cognitivism or non-cognitivism?" and "Moral motivation: internalism or externalism?". Analysing the data[115], I have selected the most relevant data for my research and have put it into a table.

I want to make use of what the survey exposes to support some of my decisions and claims. I am aware that the survey results are not exhaustively representative of all the views that philosophers hold, besides, for purposes of the survey (simplicity of categorisation and such) it remains rather unclear what hides under each of the labels ("realism", "internalism" and others). The authors of the survey explain their choice as follows: "although many of these labels are ambiguous, longer descriptions would introduce new ambiguities in turn" (Bourget and Chalmers 2013: 7-8).

The more so, the fact that majority of philosophers hold one or another view does not make that view true or preferable to others. Having acknowledged that, one can still use some of the numbers and tendencies for purposes of discerning the central debates, the dominant positions and related issues. As Bourget and Chalmers, who conducted the survey, note, "it is inevitable that some views are presupposed, other views are the focus of

[115] The most exhaustive source of the data is the website http://philpapers.org/surveys, because the paper presents the results and the limitations of the research, but not all of the data.

attention and argument, while still others are ignored" and that, therefore, some of them are used as premises in the arguments, and the rejection thereof requires argument, whereas it is the assertion of others that requires considerable justification (Bourget and Chalmers 2013: 2).

Based on the considerations of Bourget and Chalmers, the "received wisdom" is determined by what "most people believe most people believe". In order to know what passes for the "received wisdom" they have conducted a meta-survey as well: they have asked same respondents what percentage of philosophers, in their opinion, support each of the positions. As the meta-survey showed, usually, though, the respondents either overestimated or underestimated the popularity of the various philosophical positions. In this sense, the survey supposedly corrects those sociological beliefs. It is not apparent to me that the difference of those several per cent that exists between what people think to be the case and what is the *case according to the survey* indicate an error in philosophers' sociological beliefs (surveys do not reach all philosophers nor only those the most significant in the debate, besides, it is doubtful that all respondents share the same conceptions of the concepts they assign as labels to themselves and to others, etc.). However, one can agree that the survey is by and large representative of the philosophical situation. Therefore, we can base some of our decisions on it.

As the authors of the paper and survey note, knowing what deserves more argumentation and attention may improve one's work. However, I will use the results for such purposes: to locate and choose to tackle the most contentious areas and the most important debates, and to support my claims as to the dominating positions, the lability of some of the distinctions and the type of the analysis needed to settle the questions. To be more exact, it does not dictate my decisions, but, combined with the preference of basing one's theoretical inquiries on the common-sense view, it motivates and justifies a certain structure of my work.

The table of results.

		All respondents (931)		Specialists of meta-ethics (102)	
		Number of people	Percentage	Number of people	Percentage
Moral realism	accept	300	32.2	42	41.2
	lean toward	225	24.2	15	14.7
	total	525	**56.4**	57	**55.9**
Moral anti-realism	accept	123	13.2	17	16.7
	lean toward	135	14.5	10	9.8
	total	258	**27.7**	27	**26.5**
In between		89	**9.6**	16	**15.7**
Other		59	**6.4**	2	**2**
Cognitivism	accept	377	40.5	63	61.8
	lean toward	235	25.2	13	12.7
	total	612	**65.7**	76	**74.5**
Non-cognitivism	accept	53	5.7	6	5.9
	lean toward	105	11.3	8	7.8
	total	158	**17**	14	**13.7**
In between		85	**9.1**	12	**11.8**
Other		76	**8.1**	-	**-**
Internalism	accept	120	12.9	27	26.5
	lean toward	205	22	18	17.6
	total	325	**34.9**	45	**44.1**
Externalism	accept	123	13.2	23	22.5
	lean toward	154	16.5	14	13.7
	total	277	**29.7**	37	**36.2**
In between		112	**12**	16	**15.7**
Other		217	**23.2**	4	**3.9**

A note: the sum of per cents in each category does not always amount to 100; the error is paltry, though, and is due to the rounding off of some values. This table was made using the results of the aforementioned survey by Bourget and Chalmers.

Explanation and data analysis. In categorising the answers, I maintained the categories reflecting the answers, except for my categories "in between" and "other". "In between" summarises cases of rejection of the posed dichotomy, i.e. the answers such as "The question is too unclear to answer," "Accept another alternative," "Accept an intermediate view," "Accept both," "There is no fact of the matter," "Reject both". "Other" is to indicate a person's avoidance to answer due to incompetence or just one's ignoring the question: "Agnostic/undecided", "Insufficiently familiar with the issue", "Skip", "Other" (the latter is too murky to be understood as a position). Also one should keep in mind that "All respondents" include the specialists of meta-ethics as well as philosophers who specialise in other fields of philosophy; whereas another column represents only the numbers of meta-ethicists holding the various specified positions.

One can notice that as far as it concerns the question of moral realism, the percentage of answers for and against it are extremely similar among the specialists and the non-specialists of meta-ethics. Specialists usually are better informed than the non-specialists when they accept or reject a position and that can be seen in the table: relatively more specialists answer with "accept" or "reject" rather than "lean towards", also, there are fewer undecided ones ("other") among them and more of those that are likely to draw finer distinctions between the positions traditionally represented by a dichotomy of realism/anti-realism ("in between"). Whatever the respondents from both categories understand as "moral realism", it is more than twice as popular as the anti-realism. As moral realism is not defined, it means that anti-realism may contain not only the non-cognitivist views, but also the variously motivated versions of moral cognitivism.

The latter remark is supported by the data on the acceptance of cognitivism/non-cognitivism. Cognitivism is usually a necessary condition for realism, so it is of no surprise that the number of cognitivists should not be

smaller than that of realists[116]. But as the numbers show, there are more cognitivists than there are realists both among the specialists and the non-specialists. That means that those 87 cognitivist respondents, 19 of whom are specialists, adhere to either anti-realism, or to some other version of not robustly realist position, or are undecided. It may be that part of them, for example, accept some kind of error theory (cognitivism plus moral nihilism), some are constructivists (though some of constructivists may also place themselves among those who label themselves as moral realists, because we do not know how they define realism) and some possibly have a different position still.

As to the question of cognitivism/non-cognitivism, the figures differentiate dramatically: among all the respondents there are almost 4 times more philosophers who believe that cognitivism is true than those who adhere to non-cognitivism or those who hold intermediate positions or do not wish to answer the question. Among the specialists, it is even more apparent: there are more than 5 times more of those who accept or lean towards cognitivism than those who feel the same with regard to non-cognitivism, and even more than 6 times more than those who accept an intermediate position. This is very different from the realism/anti-realism case, and it is possible to make a conclusion that non-cognitivist position is marginal to the debate. The realism/anti-realism question is more of a debate than the cognitivism/non-cognitivism, especially given the relation of the former to the dominant position of cognitivism. It is central to figure out how the dominant semantic theory, and that is cognitivism, should be realised in order to fit with one or

[116] I say that *usually* cognitivism is necessary for realism, but I acknowledge that in the case of the survey it may be that some moral realist respondents accepted or leaned towards non-cognitivism. It may be that some of the respondents (especially having in mind that many of them are not meta-ethicists) do not know what the two positions exactly involve or they understand the two positions in a particular, non-orthodox, way. Also, I do not exclude the possibility that there may be such meta-ethicists who combine the two views, even if they would not be typical. But I hold that even in case of such deviations the claim that *usually* cognitivism is required for moral realism (understood in one of the two predominant meanings in moral philosophy) stands.

another ontology and moral psychology: what is it compatible with? One can also note that even among the *specialists* there are more of those who lean towards non-cognitivism than those who accept it, which indicates that there is relatively more insecurity in answering this question to the benefit of non-cognitivism in comparison to the acceptance of other positions.

As to the debate of motivational internalism/externalism, we have a different view still. Internalists and externalists receive similar number of instances of approval, even if with a slight preponderance of the internalists. It is a serious debate in which neither of the positions can be as easily dismissed as non-cognitivism in the cognitivism/non-cognitivism case. That holds even in view of the fact that part of the internalists are non-cognitivists: as we saw, non-cognitivists make only a very small part of moral philosophers, so even if all 14 who accept non-cognitivism (from those who participated in this survey) were internalists, they would still make only 1/3 of the moral philosophers who accept internalism.

Another important thing to notice in relation to the discussion of internalism/externalism is the great number of respondents who wish to skip the question (insufficiently familiar, agnostic/undecided, skip, other) along with those who take an intermediate position. 35.2 per cent of all respondents take one of these latter positions, whereas 34.9 are internalists and 29.7 externalists (also, more "lean towards" either rather than "accept" either). But it is, as can be expected, quite different among the specialists: more specialists *accept* one of the two positions in question rather than lean towards one; the percentage of "other" is negligible, whereas the proportion of those accepting an in between position is much greater than among the non-specialists. What this great proportion of neither clear-cut internalists, nor clear-cut externalists shows, I think, is that this debate is rather specialised, fine-grained – at least in its current philosophical form.

I want to draw attention that the internalism/externalism debate gained its more intersubjective meaning of "motivational internalism"/"motivational externalism" not so long ago. Some three and even two decades ago there was

much more confusion as to what claims one had in mind, what was the object of the position (is it reasons? If so, which kind of reasons – motivational or normative ones? What those reasons are – psychological or logical entities? Or perhaps it is moral judgements that we talk about?[117]). Nowadays the questions defining the debate are more or less clear, various theorists defend versions of motivational internalism and externalism knowing clearly where they belong[118].

Implications of the analysis for the structure of this research. As the survey results show, it is cognitivism about moral judgements that deserves the title of "received wisdom". This position is also supported by our common practices, by the common sense. Thus, I will concentrate on the positions that are based on this supposition not paying much attention to provision of counterarguments against non-cognitivism. After all, one always takes something for granted and tries to answer the worries that are or seem more pressing. And the question that seems to be more urgent is: in virtue of what do moral judgements have truth values? This question leads to considerations of moral ontology – not from necessity, but because the most frequent answers posit truth makers in the realm that is the object of moral ontology. We have seen that many philosophers adhere to moral realism. However, it is apparent that anti-realism is also compatible with cognitivism, and it is not obvious that moral realism is the best partner to cognitivism. Here, one could try to see if any of the two possibilities is more often chosen, but that proves to be too difficult. What complicates counting the proportion of realist cognitivism and anti-realist cognitivism among all the respondents or specialists is the absence of a definition of realism (both in the survey and in the debate: we can see what the prevalent usage is, but still there is no one universally accepted definition),

[117] A good example of these debates and varied terminology are, e.g. Williams (1981a) and Darwall (1992).

[118] Today a question that becomes more frequently asked is whether internalism should be treated as a conceptual or as an *a posteriori* truth. But I will not tackle it; instead, I will rather traditionally hold it to be a conceptual truth.

which obfuscates ascriptions. Common sense, I am afraid, is silent on the matter: one can argue at best that from a common-sense point of view morality is real, it exists or it works. But "moral realism" in most cases amounts to much more than that. So I will not exclude consideration of the anti-realist position from my work (except for the antirealist-non-cognitivist duo).

What concerns the internalism/externalism debate, it is usually said that the common-sense view, derived from our everyday practices, is internalist. But it is not as easy to discard externalism as so many meta-ethicists consider it a viable alternative. However, the need of externalism in connection to cognitivism is usually determined by one's choice to account for the truth values of moral claims by their correspondence to the state of moral affairs, thus, by (ontological) moral realism. Once (ontological) moral realism is ruled out as the best pairing to cognitivism, the reason for choosing externalism is usually gone with it.

On the other hand, as already mentioned, the internalism/externalism debate is truly nuanced and specialised, so one would need a separate voluminous work in order to do justice to it. Therefore, I do not give the debate as much attention as it deserves, neither I do it with regard to all varieties of internalism. Instead, treating externalism as a worthy opposing view, I pay it only as much attention as is necessary to neutralise its apparently harmful arguments against the most promising version of internalism.

The grain of analysis in this latter debate has to be fine – much finer than in the realism/antirealism debate. Here, as the saying goes, the devil is in the details, it is very technical, which is why the part treating internalism differs from the preceding ones in scale.

The lability of the dichotomies is the clearest from the internalism/externalism data, it is closely followed by the realism/antirealism, and cognitivism/non-cognitivism dichotomy deserves the least doubts. We can see why in the dissertation.

Thus, two effects follow. One is that cognitivism being the "received wisdom" as a semantic position, I give its two realisations – moral realism and

moral anti-realism (which I rename and use indexes instead – MR_{MI} and MR_{MD}) – attention. After giving the reasons to accept one of them rather than another, I turn to internalism in connection to cognitivism as I show that cognitivist externalism is not needed when we rule out the MR_{MI}. Two, in the second part the grain of analysis being finer, the discussion is more technical than that of the first part.

Bibliography

Adler, J. E. 2002. *Belief's Own Ethics*. Cambridge: Massachusetts Institute of Technology.

Almeida, de, C. 2001. "What Moore's Paradox Is About", *Philosophy and Phenomenological Research*, Vol. 62, No. 1 (Jan.), p. 33–58.

Aristotle. 1960. *Topica* in *Posterior Analytics. Topica*. (*Topica*), transl. by E. S. Forster. Cambridge: Harvard University Press, p. 265-739.

Baranova, J. 2004. *XX amžiaus moralės filosofija: pokalbis su Kantu*. Vilnius: VPU leidykla.

Björklund, F., Björnsson, G., Eriksson, J., Francén Olinder, R., Strandberg, C. 2012. "Recent Work: Motivational Internalism", *Analysis*, Vol. 72, No. 1, p. 124-137.

Blackburn, S. 1985. "Errors and the Phenomenology of Value" in *Morality and Objectivity. A Tribute to J. L. Mackie*, T. Honderich, ed. London: Routledge and Kegan Paul, p. 1-22.

Blackburn, S. 1987. "How to Be an Ethical Antirealist" in *Midwest Studies in Philosophy*, Vol. XII *Realism and Anti-Realism*, P. A. French, Th. E Uehling, Jr. and H. K. Wettstein, eds. Minneapolis: University of Minnesota Press, p. 361-375.

Blackburn, S. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.

Blackburn, S. 1998. *Ruling Passions*. New York: Oxford University Press.

Blackburn, S. 2010. *Practical Tortoise Raising: and other philosophical essays*. New York: Oxford University Press.

Bourget, D. and Chalmers, D. J. 2013. “What Do Philosophers Believe?”, *Philosophical Studies*, p. 1-37. doi: 10.1007/s11098-013-0259-7.

Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brink, D. O. 1997. “Moral Motivation”, *Ethics*, Vol. 108, No. 1 (Oct.), p. 4-32.

Broom, J. 2009. ”Motivation”, *Theoria*, Vol. 75, Issue 2, p. 79-99.

Broom, J. 2010. “Rationality” in *A Companion to the Philosophy of Action*, T. O’Connor, C. Sandis, eds. Singapore: Blackwell Publishing, p. 285-292.

Broome, J., Parfit, D. 1997. “Reasons and Motivation”, *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 71, p. 99-146 (Parfit: 99-130 and Broom: 131-146).

Cholbi, M. 2009. “Moore’s Paradox and Moral Motivation”, *Ethical Theory and Moral Practice*, Vol. 12, No. 5, p. 495-510.

Cholbi, M. 2011. “Depression, Listlessness, and Moral Motivation”, *Ratio*, Vol. 24, No. 1, p. 28-45.

Chisholm. R. 1977/1966. *Theory of Knowledge*, 2nd edition. Englewood Cliffs: Prentice-Hall.

Chisholm, R. 2001. “The Problem of the Criterion”, *Analytic Philosophy. An Anthology*, Martinich, A. P. and Sosa, eds. D. Oxford: Blackwell Publishers, p. 190-198.

Coates, A. 2011. “The Enkratic Requirement”, *European Journal of Philosophy*, Vol. 21, Issue 2, p. 320-333.

Dancy, J. 1993. *Moral Reasons*. Oxford: Blackwell.

Dancy, J. 2000. *Practical Reality*. Oxford: Clarendon Press.

Dancy, J. 2004. *Ethics Without Principles*. Oxford: Clarendon Press.

Darwall, S. L. 1992. "Internalism and Agency", *Philosophical Perspectives*, Vol. 6 (*Ethics*), p. 155-174.

Darwall, S., Gibbard, A., Railton, P. 1992. "Toward *Fin de siècle* Ethics: Some Trends", *Philosophy in Review: Essays on Contemporary Philosophy*, Vol. 101, No. 1, p. 115-189.

Davidson, D. 2001. "How Is Weakness of the Will Possible?" in *Essays on Actions and Events*. New York: Oxford University Press, p. 21-42.

Davidson, D. 2004a. "Incoherence and Irrationality" in *Problems of Rationality*. New York: Oxford University Press, p. 189-198.

Davidson, D. 2004b. "Paradoxes of Irrationality" in *Problems of Rationality*. New York: Oxford University Press, p. 169-188.

DePaul, M. R. 2009. "Pyrrhonian Moral Skepticism and the Problem of the Criterion", *Philosophical Issues*, Vol. 19 (*Metaethics*), p. 38-56.

DePaul, M. R. 1988. "The Problem of the Criterion and Coherence Methods in Ethics", *Canadian Journal of Philosophy*, Vol. 18, No. 1 (Mar.), p. 67-86.

Dreier, J. 1990. "Internalism and Speaker Relativism", *Ethics*, Vol. 101, No. 1 (Oct.), p. 6-26.

Dworkin, R. 1996. "Objectivity and Truth: You'd Better Believe it", *Philosophy and Public Affairs*, Vol. 25, No. 2, p. 87-139.

Foot, Ph. 1978. *Virtues and Vices*. Oxford: Blackwell.

Frankfurt, H. G. 2006. *Taking Ourselves Seriously and Getting It Right*. Stanford, California: Stanford University Press.

Gert, J. 2004. *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press.

Gert, J. 2008. "Michael Smith and the Rationality of Immoral Action", *The Journal of Ethics,* Vol. 12, No. 1, p. 1-23.

Hare, R. M. 1985. "Ontology in Ethics" in *Morality and Objectivity. A Tribute to J. L. Mackie,* T. Honderich, ed. London: Routledge and Kegan Paul, p. 39-53.

Hare, R. M. 1991/1952. *The Language of Morals*. Oxford: Oxford University Press.

Harman, G. 2002. "Internal Critique: A Logic Is Not a Theory of Reasoning and a Theory of Reasoning Is Not Logic", *Handbook of the Logic of Argument and Inference. The Turn Towards the Practical,* Vol. 1, Gabbay, D. M., Johnson, R. H., Ohlbach, H. J., Woods, J., eds. Amsterdam: Elsevier, p. 171-186.

Horgan, T., Timmons, M. 2006/1992. "Troubles for New Wave Moral Semantics: the Open Question Argument Revived" in *Arguing about Metaethics,* A. Fisher, S. Kirchin, eds. London and New York: Routledge, p. 179-199.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Jackson, F., Oppy, G. and Smith, M. 1994. "Minimalism and Truth Aptness", *Mind,* Vol. 103, Issue 411, p. 287–302.

Jokubaitis, A. 2013. „Metaetika kaip romantizmo forma“, *Problemos*, Nr. 83, p. 86-95.

Kennett, J., Smith, M. 1994. “Philosophy and Common Sense: The Case of Weakness of Will” in *Philosophy in Mind*, M. Michael, J. O'Leary Hawthorne, eds. Dordrecht: Kluwer Press, p. 141-57.

Kennett, J., Smith, M. 2004. “Frog and Toad Lose Control” in Smith, M. *Ethics and the A Priori. Selected Essays on Moral Psychology and Meta-ethics*. New York: Cambridge University Press, p. 73-83. Reprinted from *Analysis*, Vol. 56, 1996, p. 63–73.

Korsgaard, Ch. M. 1986. “Skepticism about Practical Reason“, *The Journal of Philosophy*, Vol. 83, No. 1, p. 5–25.

Korsgaard, Ch. M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, Ch. M. 2008a. “Acting for a Reason” in *The Constitution of Agency*. New York: Oxford University Press, p. 207-229.

Korsgaard, Ch. M. 2008b. “Aristotle on Function and Virtue” in *The Constitution of Agency*. New York: Oxford University Press, p. 154-173. Reprinted from the *History of Philosophy Quarterly*, Vol. 3 No. 3 (Jul. 1986), p. 259–279.

Korsgaard, Ch. M. 2008c. “Aristotle’s Function Argument” in *The Constitution of Agency*. New York: Oxford University Press, p. 129-150.

Korsgaard, Ch. M. 2008d. “From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action” in *The Constitution of Agency*. New York: Oxford University Press, p. 174-206. Reprinted from *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, S. Engstrom and J. Whiting, eds. New York: Cambridge University Press, 1996, p. 203-236.

Korsgaard, Ch. M. 2008e. “Introduction” in *The Constitution of Agency*. New York: Oxford University Press, p. 1-23.

Korsgaard, Ch. M. 2008f. “Realism and Constructivism in Twentieth-Century Moral Philosophy” in *The Constitution of Agency*. New York: Oxford University Press, p. 302-326. Reprinted from *Philosophy in America at the Turn of the Century*. APA Centennial Supplement to *The Journal of Philosophical Research*. Charlottesville, Virginia: The Philosophy Documentation Center, 2003, p. 99-122.

Korsgaard, Ch. M. 2008g. “Self-Constitution in the Ethics of Plato and Kant” in *The Constitution of Agency*. New York: Oxford University Press, p. 100-126. Reprinted from *The Journal of Ethics*, Vol. 3 (1999), p. 1-29.

Korsgaard, Ch. M. 2008h. “The Normativity of Instrumental Reason” in *The Constitution of Agency*. Oxford University Press, p. 27-68. Republished from *Ethics and Practical Reason*, G. Cullity, G. and B. Gaut, eds. Oxford: Clarendon Press, 1997, p. 213-254.

Korsgaard, Ch. M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Kuzmickas, B. 1989. „Dabartinės etikos bruožai“ in *Gėrio kontūrai: iš XX a. užsienio etikos*, B. Kuzmickas et al., sud. Vilnius: Mintis, p. 7-23.

Lenman, J. 1999. “The Externalist and the Amoralist”, *Philosophia*, Vol. 27, No. 3-4 (Nov.), p. 441-457.

Mason, E. 2008. “An Argument Against Motivational Internalism”, *Proceedings of the Aristotelian Society,* Vol. 108, Part 2, p. 135-156.

Mele, A. “Internalist Moral Cognitivism and Listlessness”, *Ethics*, Vol. 106, No. 4, p. 727-753.

Mele, A. 2003. *Motivation and Agency*. New York: Oxford University Press.

Moore, G. E. 2002/1959. "A Defence of Common-Sense" in *Philosophical Papers*. London: Routledge, p. 32-59. Reprinted from *Contemporary British Philosophy* (2nd series), J. H. Muirhead, ed., New York: Routledge, 1925.

Nagel, Th. 1986. *The View from Nowhere*. New York: Oxford University Press.

Nichols, Sh. 2002. "How Psychopaths Threaten Moral Rationalism, or Is it Irrational to Be Amoral?", *The Monist*, Vol. 85, p. 285–303.

Parfit, D. 1984. *Reasons and Persons*. New York: Oxford University Press.

Parfit, D. 2011a. *On What Matters,* Vol. 1. New York: Oxford University Press.

Parfit, D. 2011b. On What Matters, Vol. 2. New York: Oxford University Press.

Pascal, B. 1999. *Pensees* in *Pensees and other writings*, transl. by H. Levi. New York: Oxford University Press, p. 1-181.

Patapas, A. 2001. *Moralės objektyvumo problema G. Moore'o, J. Mackie ir J. Rawlso filosofijoje*, daktaro disertacija: humanitariniai mokslai, filosofija (01 H). Vilniaus universitetas.

Pettit, Ph., Smith, M. 1990. "Backgrounding Desire", *The Philosophical Review*, Vol. 99, No. 4 (Oct.), p. 565-592.

Pettit, Ph., Smith, M. 1993. "Practical Unreason", *Mind, New Series*, Vol. 102, No. 405 (Jan.), p. 53-79.

Pettit, Ph., Smith, M. 1996. "Freedom in Belief and Desire", *The Journal of Philosophy*, Vol. 93, No. 9 (Sep.), p. 429-449.

Putnam, H. 2004. *Ethics without Ontology*. Cambridge (Mass.): Harvard University Press.

Rosati, C. S. 1995. "Naturalism, Normativity, and the Open Question Argument", *Noûs*, Vol. 29, No. 1 (Mar.), p. 46-70.

Rosati, C. S. 1996. "Internalism and the Good for a Person", *Ethics*, Vol. 106, No. 2 (Jan.), p. 297-326.

Rosati, C. S. 2003. "Agency and the Open Question Argument", *Ethics*, Vol. 113, No. 3, *Centenary Symposium on G. E. Moore's Principia Ethica* (Apr.), p. 490-527.

Sayre-McCord, G. 2012. "Metaethics", *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), E. N. Zalta, ed. [Online] Available from: http://plato.stanford.edu/archives/spr2012/entries/metaethics/.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell Publishing.

Smith, M. 1995. "Internal Reasons", *Philosophy and Phenomenological Research*, Vol. 55, No. 1 (Mar.), p. 109-131.

Smith, M. 1996. "Normative Reasons and Full Rationality: Reply to Swanton", *Analysis*, Vol. 56, No. 3 (Jul.), p. 160-168.

Smith, M. 1997. "In Defence of 'The Moral Problem': A Reply to Brink, Copp, and Sayre-McCord", *Ethics*, Vol. 108, No. 1 (Oct.), p. 84-119.

Smith, M. 2001. "The Incoherence Argument: Reply to Schafer-Landau", *Analysis*, Vol. 61, No. 3 (Jul.), p. 254-266.

Smith, M. 2002a. "Bernard Gert's Complex Hybrid Conception of Rationality" in *Rationality, rules, and ideals: critical essays on Bernard Gert's moral theory*, W. Sinnott-Armstrong, R. Audi, eds. Lanham: Rowman & Littlefield, p. 109-124.

Smith, M. 2002b. “Evaluation, Uncertainty and Motivation”, *Ethical Theory and Moral Practice*, Vol. 5, p. 305-320.

Smith, M. 2004a. “A Theory of Freedom and Responsibility” in *Ethics and the A Priori*. *Selected Essays on Moral Psychology*. Cambridge: Cambridge University Press, p. 84-113. Reprinted from *Ethics and Practical Reason*, G. Cullity, B. Gaut, eds. Oxford: Oxford University Press, 1997, p. 293-319.

Smith, M. 2004b. “Instrumental Desires, Instrumental Rationality”, *Proceedings of the Aristotelian Society, Supplementary Volumes,* Vol. 78, p. 93-109.

Smith, M. 2004c. “Moral Realism” in *Ethics and the A Priori*. *Selected Essays on Moral Psychology*. Cambridge University Press, p. 181-207. Reprinted from *Blackwell Guide to Ethical Theory*, H. LaFollette, ed. Oxford: Blackwell, 2000, p. 15-37.

Smith, M. 2007. “Is there a Nexus between Reasons and Rationality” in *Moral Psychology (Poznań Studies in the Philosophy of the Sciences and the Humanities)*, Vol. 94, S. Tenenbaum, ed. Amsterdam and New York: Rodopi, p. 279-298.

Smith, M. 2009. “Reasons with Rationalism after All”, *Analysis Reviews*, Vol. 69, No. 3 (Jul.), p. 521-530.

Stocker, M. 1979. ”Desiring the Bad”, *Journal of Philosophy*, Vol. 76, p. 738-753.

Strandberg, C. 2012a. “A Dual Aspect Account of Moral Language”, *Philosophy and Phenomenological Research*, Vol. LXXXIV (84), No. 1 (Jan.), p. 87-122.

Strandberg, C. 2012b. “An Internalist Dilemma – and an Externalist Solution”, *Journal of Moral Philosophy*, Vol. 10 Issue 1, p. 25-51.

Strandberg, C., Björklund, F. 2013. “Is Moral Internalism Supported by Folk Intuitions?”, *Philosophical Psychology*, Vol. 26, Issue 3, p. 319-335.

Street, Sh. *Forthcoming*. “Objectivity and Truth: You’d Better Rethink It”, *Philosophy & Public Affairs*. P. 1-44. [Online] Available from: https://files.nyu.edu/ss194/public/sharonstreet/Writing_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You%27d%20Better%20Rethink%20It.pdf.

Street, Sh. 2010. “What is Constructivism in Ethics and Metaethics?”, *Philosophy Compass*, Vol. 5, Issue 5, p. 363–384.

Sturgeon, N. L. 2006/1985. “Moral Explanations” in *Arguing about Metaethics*, A. Fisher, S. Kirchin, eds. London and New York: Routledge, p. 117-144.

Svavarsdóttir, S. 1999. “Moral Cognitivism and Motivation”, *The Philosophical Review*, Vol. 108, No. 2 (Apr.), p. 161-219.

Svavarsdóttir, S. 2006. “How Do Moral Judgments Motivate”, in *Contemporary Debates in Moral Theory*, J. Dreier, ed. Malden (Mass.): Blackwell Publishing, p. 163-181.

Tresan, J. 2009. “The Challenge of Communal Internalism”, *The Journal of Value Inquiry*, Vol. 43, p. 179-199.

Vasilionytė, I. 2012. “The Open Question Argument and the Possibility of Reductionism”, *Problemos, Supplement*, p. 37-50.

Wedgwood, R. 2002. “Practical Reason and Desire”, *Australasian Journal of Philosophy*, Vol. 80, No. 3 (Sept.), p. 345-358.

Williams, B. 1981a. “Internal and External Reasons” in *Moral Luck. Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press,

p. 101-113. Reprinted from *Rational Action. Studies in Philosophy and Social Science*, R. Harrison, ed. Cambridge: Cambridge University Press, 1979, p. 17-28.

Williams, B. 1981b. "Persons, Character and Morality" in *Moral Luck. Philosophical Papers 1973-1980*. Cambridge University Press, p. 1-19. Reprinted from *The Identities of Persons*, A. O. Rorty, ed. Berkley and Los Angeles: University of California Press, 1976, p. 197-216.

Williams, B. 1985. "Ethics and the Fabric of the World" in *Morality and Objectivity. A Tribute to J. L. Mackie*, T. Honderich, ed. London: Routledge and Kegan Paul, p. 203-214.

Williams, B. 2006/1985. *Ethics and the Limits of Philosophy*. London and New York: Routledge.

Yaffe, G. and Nichols, R. 2009. "Thomas Reid", *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), E. N. Zalta, ed. [Online] Available from: http://plato.stanford.edu/archives/win2009/entries/reid/.

Zangwill, N. 2005. "Moore, Morality, Supervenience, Essence, Epistemology", *American Philosophical Quarterly*, Vol. 42, No. 2 (Apr.), p. 125-130.

Zangwill, N. 2008. "The Indifference Argument", *Philosophical Studies*, Vol. 138, No. 1, p. 91-124.

Zangwill, N. 2012. "Rationality and Moral Realism", *Ratio* (New series), Vol. 25 No. 3, p. 345-364.